



Critical Care Reviews

Book 2026

The Best Critical Care Trials of 2025

First Edition





**Belfast Health and
Social Care Trust**

Critical Care Charitable Trust Fund

Rob Mac Sweeney's work with Critical Care Reviews has been
generously supported by the Critical Care Charitable Trust Fund
of the Belfast Health and Social Care Trust

About

The Critical Care Reviews Book 2026 offers a closer look at the clinical research that emerged during 2025. Faced with a steady flow of new trials, clinicians need more than a headline result. They need to understand how a study was conducted, what its findings support, and where uncertainty remains.

This volume approaches those questions through detailed accounts of selected trials first published online that year. The chapters examine the reasoning behind each study, the choices made in its design and analysis, and the implications of its results. Earlier research provides the context, with the discussion reflecting evidence available by 31 December 2025.

None of this evidence would exist without the patients and families who take part in research, or the clinical and research teams who deliver it. We owe them our thanks. Their efforts matter whether an intervention proves helpful, causes harm or fails to improve the outcome it was intended to change.

Access to that knowledge should extend beyond those able to pay for it. The book is therefore shared freely, with grateful acknowledgement of the generous support provided by the Critical Care Charitable Trust Fund of the Belfast Health and Social Care Trust. Its contribution helps us bring independent critical care education to colleagues wherever they work.

No annual collection can include every worthwhile study, and our interpretations will not command universal agreement. We welcome that diversity of opinion. Read alongside the original papers, we hope these appraisals will prompt discussion, sharpen clinical judgement and help readers decide which findings deserve a place in their own practice.

Rob Mac Sweeney
Critical Care Reviews
Belfast

Disclaimer

This book aims to summarise major critical care trials of 2025. Although care has been taken to ensure information is correct, this is not guaranteed and no responsibility is accepted for clinical decisions based on material within this book. Clinicians are advised to check the primary literature at all times. The opinions stated within this book do not constitute clinical advice. They are opinions, not fact, and others may take a different view of our interpretations of these trials. Please refer to the appropriate clinical guideline issued by the relevant society or scientific body for the management of any specific condition.

This book was written with the assistance of AI.

Table of Contents

■ Neuro Trials	1
NEUROTHERM	2
4D	11
DISTAL	20
BRIDGE-TNK	26
ESCAPE-MeVO	34
PEARL	42
A2B	50
B-Free	58
MIND	67
POST-TNK	76
POST-UK	85
■ Circulatory Trials	95
Altshock-2	96
LEVOECMO	104
EVERDAC	112
FLUID	120
STORM-PE	128
CELEBRATE	137
IMPROVE-multi	146
PRINCE	155
■ Airway Trials	164
RSI	165
PreVent 2	173
■ Respiratory Trials	181
SESAR	182
UK-ROX	191
PROMIZING	199
ACTiVE	207
RESCUE-3	216
DESIGNATION	225
NAVIGATE	234

Table of Contents

Gastrointestinal Trials	243
TIMING	244
Empirical antifungals in ACLF	253
Nutrition Trials	262
TARGET Protein	263
Renal Trials	272
BICARICU-2	273
LIBERATE-D	282
BigpAK-2	291
ALBICS-AKI	300
Haematology Trials	309
FARES-II	310
ANH	320
TOP	329
CLIP-II	338
Sepsis Trials	347
OPTPRESS	348
ANDROMEDA-SHOCK-2	356
TARTARE-2S	365
PIC	374
SuDDICU	383
ImmunoSep	392
SONIA	401
REMAP-CAP: Corticosteroids	410
ENCOMPASS	419
INHALE-WP3	428
Paediatric Trials	437
GASTROSAM	438
IsoCOMFORT	447
Kids THRIVE	457
REDvent	466
TORPIDO 30/60	476



Neuro Trials

NEUROTHERM

de Mesmay M, Gera L, Gregoire C, Roy M, Welschbillig S, Le Cossec C, et al. Effect of Paracetamol on Cerebral Temperature in Febrile Brain-Injured Patients. The NEUROTHERM Study: A Randomized Controlled Pharmacodynamic Trial. Crit Care Med. 2026;54(1):108-118.

Introduction

Fever after acute brain injury is associated with worse outcomes and increases concern about secondary cerebral injury. Treating it is common practice, but three questions should be distinguished: whether a treatment lowers measured temperature, whether it controls temperature in brain tissue, and whether the resulting change improves neurological recovery. Evidence for one does not necessarily establish the others. The distinction is particularly relevant to paracetamol, a familiar first-line treatment whose apparent simplicity can obscure uncertainty about its effects in severely injured brains.

Most antipyretic studies measure systemic temperature. An intraparenchymal probe can reveal a different temperature from a bladder, rectal or other peripheral measurement. A patient may therefore meet a systemic temperature target while brain tissue remains warmer. Conversely, inserting an invasive monitor merely to obtain a more direct measurement requires an outcome justification beyond demonstration that two measurement sites differ.

Paracetamol has produced modest systemic temperature reductions in previous critical care trials. A small study of repeated dosing after traumatic brain injury did not establish a reduction in its primary temperature endpoint, while a single-dose study in febrile general ICU patients found lower temperature accompanied by lower blood pressure. Neither directly answered the question of antipyretic efficacy within brain tissue in patients with established cerebral fever.^{1,2}

NEUROTHERM was designed to address that narrower pharmacodynamic question. It compared one intravenous dose of paracetamol with saline placebo in deeply sedated patients already undergoing invasive cerebral monitoring. Its relevance lies in defining the size, time course and variability of the cerebral temperature response, rather than demonstrating neuroprotection or establishing the best overall fever-management strategy.³

Synopsis

NEUROTHERM was a single-centre, randomised, double-blind, placebo-controlled trial conducted in the neuro-ICU of the Rothschild Foundation Hospital in Paris. Recruitment ran from May 2018 to May 2021. The hospital supported the study. Written consent was obtained from patients or relatives, and the trial was registered as NCT03648021. Although its final issue citation is dated January 2026, the principal report first appeared online in November 2025 and falls within the CCR26 publication-year scope.³

Eligible patients were adults admitted with primary vascular, traumatic or postoperative neurosurgical brain injury, expected to require more than two days of sedation, and already monitored with an intracranial-pressure sensor incorporating a thermal probe. Cerebral temperature had to be at least 38.5°C for more than 30 minutes. Where possible, treatment was studied during the first febrile episode requiring intervention. Exclusions included paracetamol allergy, severe liver disease, an intervention during the preceding six hours capable of affecting temperature, pregnancy or breastfeeding, and previous participation.

The intraparenchymal sensor was positioned approximately 2 cm beneath the cortical surface as part of usual clinical care. Systemic temperature was measured using a bladder catheter thermistor. Patients received midazolam and sufentanil, targeting a Richmond Agitation-Sedation Scale score of -4 or -5. Usual cerebral haemodynamic management targeted intracranial pressure below 22 mm Hg and cerebral perfusion pressure above 70 mm Hg, or an individualised perfusion target informed by transcranial Doppler or brain tissue oxygenation.

Computer-generated allocation used blocks of eight and assigned patients in a 1:1 ratio. Those weighing at least 50 kg received 1 g intravenous paracetamol in 100 mL over 15 minutes; the specified dose below 50 kg was 15 mg/kg. The placebo group received 100 mL saline over the same period. Preparations appeared identical, and treatment was masked to the nurse and physician caring for the patient. The study evaluated a single administration, not a course of repeated antipyretic treatment.

Cerebral and systemic temperature, heart rate, arterial pressure, intracranial pressure, cerebral perfusion pressure and catecholamine administration were recorded every ten minutes over six hours. Time zero was the start of the infusion. Clinicians could deepen sedation, administer neuromuscular blockade or use external cooling if hyperthermia was poorly tolerated. These additional measures were recorded at six hours. Thus, placebo assignment did not preclude rescue temperature control.

The primary endpoint was the between-group difference in mean cerebral temperature over the six hours after treatment. Secondary outcomes included systemic temperature, the maximum-to-minimum cerebral temperature difference, time spent below 38.5°C, the proportion ever reaching that threshold, additional temperature-control treatment and haemodynamic changes. These endpoints describe antipyretic action and physiological tolerance. Functional neurological recovery was not the primary purpose of the trial.

The sample-size calculation considered a 0.5°C difference clinically relevant. With 80% power and an alpha of 0.05, a mixed linear regression framework incorporating repeated measurements required 48 patients per group under the specified variance assumptions. The report describes intention-to-treat analysis, mean imputation for missing primary-endpoint data and an available-data sensitivity analysis. Continuous and categorical comparisons used the tests specified in the principal methods. An additional responder analysis was explicitly post hoc: failure to achieve cerebral temperature below 38.5°C at any point was labelled non-response.

The flow diagram shows 100 patients assessed, one excluded for not meeting eligibility criteria and 99 randomised: 49 to paracetamol and 50 to placebo. Forty-seven paracetamol patients received the full allocated intervention; all 49 remained in the primary analysis. Mean age was approximately 55 years, 24% were women, and 91% had neurovascular disease. Intracerebral haemorrhage, subarachnoid haemorrhage and ischaemic stroke predominated. Only five patients had severe head trauma, all in the placebo arm. Median time in the neuro-ICU before inclusion was three days.

Mean cerebral temperature at inclusion was 39.0°C in both groups, compared with systemic temperature of 38.6°C. Mean intracranial pressure was 9 mm Hg; only four patients had a pressure above 22 mm Hg at inclusion. Approximately half had an external ventricular drain, and 44% were being treated for infection. These were severely brain-injured, deeply sedated patients, but most were not experiencing intracranial hypertension when the study dose was given.

Mean cerebral temperature during the six-hour observation period was 38.4°C with paracetamol and 39.0°C with placebo, with standard deviations of 0.5°C in both groups and $P < 0.001$. The report describes an overall difference of approximately 0.6°C. Median time with cerebral temperature below 38.5°C was 215 minutes (interquartile range 0-290) versus zero minutes (0-5), also with $P < 0.001$. Among patients with available threshold data, 33/47 receiving paracetamol and 15/49 receiving placebo crossed below 38.5°C, or 70.2% versus 30.6%.³

Four paracetamol patients and two placebo patients required additional temperature-control measures. All six received neuromuscular blockade and three also received external cooling. In the post hoc paracetamol analysis, 14/47 patients never crossed below 38.5°C. Among those who did, median time to reaching the threshold was 70 minutes. The subgroup comparisons did not identify a convincing baseline predictor of threshold response.

Systemic temperature results were consistent with an antipyretic effect: mean six-hour temperature was 38.1°C versus 38.7°C. Across the study, cerebral temperature averaged approximately 0.3°C above systemic temperature. Mean systolic-pressure change from baseline was -9 versus -2 mm Hg, and heart-rate change was -4 versus zero beats/min. Mean arterial pressure, cerebral perfusion pressure and intracranial pressure did not show convincing between-group differences. The supplement lists substantial missingness for norepinephrine-dose comparisons. The main report records no harms or unintended effects, but the study was small and observation of the intervention's immediate effects lasted only six hours.⁴

Critique

NEUROTHERM establishes a short-term cerebral antipyretic effect of intravenous paracetamol in this selected population. It does not establish that lowering cerebral temperature by this amount prevents secondary injury or improves recovery. This is a boundary of the question, rather than evidence that the pharmacodynamic result is unimportant. Knowing whether a commonly used treatment reaches its intended physiological target is useful, particularly when the target organ is measured directly. The clinical error would be to translate a positive temperature endpoint automatically into an outcome recommendation.

The treatment effect is convincing at the group level. Randomisation, an indistinguishable placebo and masking of bedside staff protect against several explanations based on treatment expectations. Frequent simultaneous brain and bladder measurements provide a detailed response profile, and the size of the observed difference exceeded the 0.5°C difference used in planning. Nevertheless, the principal table supplies group means, standard deviations and a P value rather than a directly reported confidence interval for the between-group primary contrast. The chapter preserves that reporting instead of manufacturing an interval from incomplete assumptions about repeated observations and imputation.

The distinction between the overall treatment effect and the responder description is especially important. Approximately 30% of treated patients never crossed below

38.5°C, but approximately 31% of placebo patients did cross that threshold. A threshold-defined response therefore incorporates spontaneous temperature variation as well as drug action. Furthermore, the responder groups were constructed after treatment. Comparing their temperature trajectories cannot estimate what would happen if a clinician could prospectively select a biologically responsive subgroup. The roughly 1°C fall among selected responders should not replace the randomised comparison as the expected treatment effect.

Nor does failure to cross a threshold necessarily prove pharmacological resistance. A patient might experience a reduction while remaining above 38.5°C, and a patient beginning closer to the threshold may cross it after a smaller change. Baseline temperatures were broadly similar, but that does not validate the responder classification as a distinct phenotype. The study found no reliable bedside predictors of non-response; it was not large enough to exclude meaningful predictors. Its practical contribution is to support reassessment after treatment, rather than to supply a validated selection rule or a precise timetable for mandatory escalation.

The measurement comparison also needs qualification. A mean cerebral-to-bladder gradient of about 0.3°C demonstrates that the sites were not interchangeable in this cohort. It does not justify adding 0.3°C to every bladder measurement in every patient, because an average difference is not a patient-level calibration. Cerebral temperature is itself sampled from one intraparenchymal location, not from the whole brain. The trial did not randomise temperature-management decisions guided by brain versus systemic measurements, so it cannot show that invasive temperature monitoring improves care or justify inserting an intracranial device solely for fever management.

Patient selection is consequential. Ninety-one per cent had neurovascular disease, and no paracetamol recipient had severe traumatic brain injury. All were deeply sedated and already considered suitable for invasive monitoring under one unit's practice. Findings may differ in awake stroke patients, in predominantly traumatic cohorts, after cardiac arrest, or in patients with less severe injury. Mean intracranial pressure was low, and the authors distinguished this pharmacodynamic study from management of refractory intracranial hypertension. It would be particularly unsafe to interpret the six-hour placebo observation as a recommendation to tolerate fever during a cerebral pressure crisis.

Rescue treatment was available and used in six patients. That preserves clinical discretion but means some observed temperature trajectories include additional interventions. The small numbers do not suggest that paracetamol clearly reduced the need for rescue treatment; numerically, rescue was more common in its group. No firm

conclusion should be drawn from that imbalance. It does, however, underline the difference between lowering average temperature and reliably controlling clinically poorly tolerated fever. The trial cannot determine which patients should receive physical cooling immediately or whether a combined strategy is preferable.

The haemodynamic findings deserve attention without exaggeration. Paracetamol lowered systolic pressure and heart rate more than placebo, while mean perfusion pressure and intracranial pressure were similar. Active haemodynamic support can preserve perfusion despite a drug-related pressure tendency, and norepinephrine comparisons had missing data in 26 patients per arm according to the supplement. Absence of a significant mean difference in cerebral perfusion pressure therefore does not prove absence of haemodynamic risk. Equally, the available data do not demonstrate cerebral hypoperfusion caused by the study dose. Monitoring treatment response remains relevant in a population for whom perfusion pressure matters.

Missingness and the analytical description impose smaller but real limits. Forty-nine patients were assigned paracetamol, but the threshold-response analysis used 47; placebo threshold data were available for 49. Mean imputation in the primary analysis may understate variability, depending on how it was implemented. The report's sample-size model and its description of final statistical testing are not a complete reproducible account of longitudinal modelling. These issues should encourage precise denominator reporting and restraint about effect precision. They do not, on the presented evidence, negate the sizeable randomised temperature separation.

Finally, a single dose over six hours cannot establish the effects of repeated administration, the ideal treatment threshold, or the best route and dose across neurocritical care. Infection was being treated in many patients, but infectious and neurogenic fever were not reliably distinguished. A familiar drug may offer a useful initial antipyretic action while leaving these broader questions unresolved. NEUROTHERM supports that limited, measurable action; claims of neuroprotection, universal response or comprehensive safety would go beyond its results.

Body of Evidence

Earlier paracetamol studies largely evaluated systemic temperature or functional outcomes in populations different from NEUROTHERM. The PAIS trial enrolled 1,400 patients with acute ischaemic stroke or intracerebral haemorrhage and baseline temperature between 36°C and 39°C. Treatment was 6 g paracetamol daily or placebo, begun within 12 hours. Improvement beyond the expected modified Rankin Scale outcome occurred in 37% versus 33%, with an adjusted odds ratio of 1.20 (95% CI 0.96-1.50). The primary result did not establish clinical benefit. A favourable post hoc subgroup finding among patients starting at 37-39°C required confirmation. This historical high-dose regimen should not be mistaken for the single-dose strategy tested by NEUROTHERM.⁵

Saxena and colleagues studied 41 patients with traumatic brain injury, assigning intravenous paracetamol or saline every four hours for 72 hours. The historical regimen delivered 6 g daily. Mean systemic core temperature was 37.4°C versus 37.7°C, a difference of -0.3°C (95% CI -0.6 to 0.0; P=0.09). This was routine administration in a relatively normothermic group, rather than treatment triggered by sustained cerebral fever near 39°C. The small, inconclusive systemic-temperature result therefore does not directly contradict NEUROTHERM's positive cerebral-temperature result. Nor does either study establish improved neurological outcome from repeated paracetamol dosing.¹

The 40-patient trial by Schell-Chaple and colleagues provides a closer pharmacodynamic comparison. Febrile general ICU patients received 1 g intravenous paracetamol or placebo over 15 minutes, followed for four hours. Adjusted time-weighted temperature was 0.47°C lower with paracetamol (95% CI -0.76 to -0.18). Systolic pressure, mean arterial pressure and heart rate were also lower. Its systemic-temperature finding is consistent with a modest antipyretic effect, while its blood-pressure findings support attention to haemodynamic tolerance. NEUROTHERM extends the measurement to the brain in an invasively monitored neurocritical care population.²

INTREPID addressed whether preventing fever with a more intensive strategy improves recovery after vascular brain injury. It randomised 677 patients to automated surface temperature management targeting 37°C or standardised treatment of fever when it occurred. Fever burden was lower with prevention, but three-month functional outcome did not improve convincingly: the odds ratio for a favourable modified Rankin Scale shift was 1.09 (95% CI 0.81-1.46). Recruitment stopped after a planned futility analysis for that functional endpoint. The intervention, duration and primary measurement differed from NEUROTHERM, but the interpretative lesson is direct: effective temperature control is not itself proof of functional benefit.⁶

These studies answer complementary questions rather than forming a single interchangeable evidence pool. PAIS examined early pharmacological treatment and recovery after stroke; the traumatic-brain-injury study examined repeated systemic antipyresis; Schell-Chaple examined one dose and immediate haemodynamics; INTREPID examined prolonged device-based prevention. NEUROTHERM supplies direct evidence that a standard intravenous dose can lower measured cerebral temperature after established fever, with incomplete threshold response and a modest systolic-pressure effect.

By the end of 2025, that finding supports paracetamol as an antipyretic option whose effect must be observed, rather than assumed. It leaves unanswered whether cerebral-temperature-directed care outperforms systemic-temperature-directed care, which threshold should trigger intervention, and whether any particular temperature-control strategy improves neurological recovery in the patients studied. Those questions require trials designed around patient-centred outcomes and the consequences of the complete treatment strategy.

Does intravenous paracetamol reliably control cerebral fever after severe brain injury?

It lowers cerebral temperature on average, but control is incomplete. Reassess temperature and haemodynamics; neurological benefit remains unproven.

References

1. Saxena MK, Taylor C, Billot L, Bompont S, Gowardman J, Roberts JA, et al. The Effect of Paracetamol on Core Body Temperature in Acute Traumatic Brain Injury: A Randomised, Controlled Clinical Trial. PLoS One. 2015;10(12):e0144740.
2. Schell-Chaple HM, Liu KD, Matthay MA, Sessler DI, Puntillo KA. Effects of IV Acetaminophen on Core Body Temperature and Hemodynamic Responses in Febrile Critically Ill Adults: A Randomized Controlled Trial. Crit Care Med. 2017;45(7):1199-1207.
3. de Mesmay M, Gera L, Gregoire C, Roy M, Welschbillig S, Le Cossec C, et al. Effect of Paracetamol on Cerebral Temperature in Febrile Brain-Injured Patients. The NEUROTHERM Study: A Randomized Controlled Pharmacodynamic Trial. Crit Care Med. 2026;54(1):108-118.

4. de Mesmay M, Gera L, Gregoire C, Roy M, Welschbillig S, Le Cossec C, et al. Supplemental digital content of the NEUROTHERM study. Critical Care Medicine. Published with the principal report, online 2025.
5. den Hertog HM, van der Worp HB, van Gemert HM, Algra A, Kappelle LJ, van Gijn J, et al. The Paracetamol (Acetaminophen) In Stroke (PAIS) trial: a multicentre, randomised, placebo-controlled, phase III trial. Lancet Neurol. 2009;8(5):434-440.
6. Greer DM, Helbok R, Badjatia N, Ko SB, Guanci MM, Sheth KN. Fever Prevention in Patients With Acute Vascular Brain Injury: The INTREPID Randomized Clinical Trial. JAMA. 2024;332(18):1525-1534.

Possible Reporting Inconsistencies. The supplied report gives a six-hour time-below-threshold result for cerebral temperature below 38.5°C, but later describes the same duration using 38°C. The chapter follows Table 2's 38.5°C definition. Table 3's diagnostic subgroup counts do not reconcile with the paracetamol-arm diagnostic counts in Table 1; these post hoc diagnostic counts are not used to infer a treatment-response phenotype. The response percentages use 47 paracetamol and 49 placebo observations, rather than the full randomised denominators.

Godet T, Louis C, Rieu B, De Jong A, Couhault P, Pradel G, et al. Dexmedetomidine for treatment of hyperactive delirium in non-intubated ICU patients: the 4D randomized clinical trial. *Intensive Care Med.* 2025;51(12):2305-2317.

Introduction

Agitated delirium in a patient without a protected airway creates an immediate therapeutic dilemma. Agitation may disrupt oxygen delivery, essential treatment and invasive devices, yet increasing sedation can impair ventilation and precipitate intubation. Treating pain, hypoxia, infection, withdrawal and environmental distress remains necessary, but these measures do not always control dangerous behaviour promptly. A sedative that permits arousal and has relatively little respiratory depressant effect is therefore attractive.

Dexmedetomidine offers that pharmacological profile, although hypotension and bradycardia can limit its use. Evidence from ventilated patients with agitation preventing extubation suggested a potential role. A non-randomised study also described successful rescue treatment in non-intubated patients who had not responded to haloperidol. Neither comparison established the effect of starting dexmedetomidine in a broader non-intubated population with hyperactive delirium. More fundamentally, reducing motor agitation is not necessarily the same as resolving the underlying disorder of attention and cognition.^{1,2}

The 4D trial addressed this gap using a blinded comparison with placebo, with immediate and rescue treatment available in both groups. Its primary outcome combined time to agitation control, time to a negative delirium assessment and intubation for deep sedation. That design captured several stages in a clinically important escalation pathway. It also created an interpretative challenge: a favourable combined result could arise from improvement in one component without demonstrable improvement in the others.³

Synopsis

4D was an investigator-initiated, double-blind, randomised trial in nine French ICUs. Recruitment occurred from December 2017 to February 2022. Support came from the French Ministry of Health, Clermont-Ferrand university hospitals and the AZUREA network; the funders had no role in trial conduct, analysis or publication. Consent could be deferred until the patient or a representative could decide about continuing participation and use of collected data.^{3,4}

Adults were eligible if they were not receiving invasive ventilation, had a positive Confusion Assessment Method for the ICU assessment and a Richmond Agitation–Sedation Scale score of at least +1. Recent extubation or assisted ventilation through a tracheostomy within 24 hours excluded participation. Patients could not have received haloperidol or dexmedetomidine in the preceding 72 hours. Further exclusions included alcohol-withdrawal delirium, contraindications to either study-related drug, severe neuropsychiatric conditions, permanent inability to consent and circumstances making delirium assessment unreliable. The supplement describes initial attempts at reorientation, environmental adjustment and pain control before enrolment.

A web-based system assigned patients equally to dexmedetomidine or saline placebo, with centre stratification and blocks of two. An independent staff member prepared the infusion away from the patient's treating team. Numbered opaque envelopes concealed the different commercial containers, and those involved in treatment and assessment remained masked. The small block size made allocation concealment particularly important.

Dexmedetomidine began at 0.2–0.5 micrograms/kg/hour and could be adjusted between 0.2 and 1.4 micrograms/kg/hour, targeting a sedation score of 0 or –1. No loading bolus was described. The infusion continued for at least 36 hours after delirium resolution, defined for this purpose by four consecutive negative delirium assessments, or until ICU discharge. Patients with agitation scores of at least +2 immediately received intravenous haloperidol 2.5 mg while the study infusion was started. Further haloperidol was permitted up to 30 mg daily, with additional open-label rescue treatment at the clinician's discretion if necessary. The comparison was therefore addition of dexmedetomidine to an active management pathway, not sedation versus no treatment.

Agitation was assessed hourly during the first six hours and then every four hours. Delirium was assessed at baseline and every 12 hours. The primary outcome combined the delay to an agitation score below +1, the delay to a negative delirium assessment and the requirement for intubation with deep sedation to control delirium. Continuous and binary components were standardised, averaged and multiplied by 100. The

presented results interpret a lower combined score as favourable, although the description of score direction contains inconsistent wording. The score is a statistical summary rather than a directly interpretable number of hours, days or intubations prevented.

Secondary outcomes included the individual components, days alive without delirium or ventilation, ICU stay, mortality, infection, recurrence, rescue medication and adverse events. The initial calculation required 110 patients per group at 90% power and a 0.018 significance level; the target was increased to 150 per group to accommodate deviations and losses. A planned interim analysis after approximately 150 participants used a stopping threshold of $P < 0.01$. The main comparison used the Mann–Whitney test, with additional adjusted and per-protocol analyses. Secondary tests were exploratory and were not adjusted for multiple comparisons.

Of 344 patients screened, 168 were enrolled and randomised. Seventeen subsequently withdrew permission for data use, leaving 77 dexmedetomidine recipients and 74 controls in the population labelled intention-to-treat. Primary outcome data were complete for these 151 patients. Six further eligibility violations were excluded from the per-protocol population. The trial stopped for efficacy following its planned interim analysis, before reaching the intended 300 participants. The supplied appendix reproduces a stopping report dated September 2023, after the reported recruitment period; the publication's account of monitoring arrangements is not entirely consistent with the appendix's description.

Patients were predominantly men, with median ages of 69 and 72 years in the dexmedetomidine and placebo groups. Approximately half had medical admissions, with the remainder admitted after elective or urgent surgery. Randomisation occurred a median of four and five days after ICU admission. Median baseline agitation scores were +1 in both groups, with interquartile ranges of +1 to +2. Thus, the population included restless patients and was not restricted to severe, combative agitation.

The median infusion duration was four days in each group. The median dexmedetomidine dose was 0.32 micrograms/kg/hour; the placebo-equivalent rate was 0.46. On the inclusion day, haloperidol was given to 18 of 77 dexmedetomidine patients (23.4%) and 30 of 74 controls (40.5%). Other rescue medication was used in 16.9% and 32.4%. During the first 30 days, any open-label rescue medication was used in 58.4% and 81.1%, respectively, with a reported absolute difference of –23 percentage points (95% confidence interval –37 to –8).^{3,4}

The primary composite favoured dexmedetomidine: the reported median contrast was –30.9 rescaled points (95% confidence interval –49.4 to –12.4), with $P=0.001$, crossing the interim stopping threshold. Median time to agitation control was one hour (interquartile range 1–2) with dexmedetomidine and two hours (1–7) with placebo. The reported median difference was –1.0 hour (95% confidence interval –2.0 to –0.1).³

The other components did not show clear improvement. Time to a negative delirium assessment was approximately 0.8 versus one day, with $P=0.38$. Intubation occurred in two of 77 patients (2.6%) and three of 74 (4.1%), giving a relative risk of 0.64 (95% confidence interval 0.11–3.75). Median ICU stay after randomisation was six days in each group. Day-30 deaths were 10 of 74 patients with follow-up in the dexmedetomidine group (13.5%) and 13 of 72 controls (18.1%); the relative risk was 0.75 (95% confidence interval 0.35–1.60). Five patients lacked day-30 vital status.

Hypotension occurred in 52 of 77 dexmedetomidine patients (67.5%) and 41 of 74 controls (55.4%); treatment was required in 44.2% and 33.8%. The reported difference in treated hypotension was 10 percentage points (95% confidence interval –5 to 26). Bradycardia requiring treatment occurred in three and two patients. No serious adverse event was classified as related to study treatment. Delirium recurrence was numerically more frequent with dexmedetomidine, but uncertainty was substantial. The per-protocol and adjusted primary analyses remained favourable. There was no evidence of an age-by-treatment interaction; an exploratory metabolic-phenotype interaction was not sufficient to define a treatment subgroup.^{3,4}

Critique

4D demonstrates a signal for faster agitation control, not a general resolution of the problems associated with delirium. The positive composite is chiefly supported by its agitation component. The trial did not establish shorter delirium, fewer intubations, shorter ICU stay or improved survival. These distinctions are essential because describing the intervention simply as an effective treatment for delirium implies more than the individual outcomes support.

The composite combines events of very different clinical importance and frequency. One hour less agitation may matter greatly when a patient is removing oxygen or threatening an essential device, but it is not interchangeable with avoiding intubation. Standardisation places variables on comparable statistical scales; it does not supply patient-centred weights. Only five intubations occurred, so the study contributes little precision to the most consequential escalation outcome. The 30.9-point difference is neither a percentage improvement nor a validated measure of clinical benefit.³

The median agitation difference also needs context. Assessments were hourly initially, putting the one-hour median in the active group at the first scheduled reassessment. That limits temporal resolution. The placebo distribution had a longer upper tail, so focusing only on the medians may understate the experience of prolonged agitation in some controls. However, post hoc comparisons at several thresholds cannot independently establish which patients benefit most. Both the overall distribution and the small absolute median improvement should inform interpretation.

The intervention was resource-intensive relative to that immediate result. Infusions continued for a median of four days, including treatment after sustained delirium resolution. The study did not compare that strategy with a brief infusion stopped once agitation settled. A clinician weighing the result must therefore consider the whole treatment exposure, monitoring and haemodynamic effects, rather than equate one hour faster control with one hour of drug administration. The optimum duration remains unanswered.

The comparator was clinically active and ethically sensible. Patients with more marked agitation could receive haloperidol immediately, and both groups could receive further rescue medication. This allowed treatment of distress or danger while retaining a masked randomised comparison. Differences in subsequent rescue use may be part of dexmedetomidine's effect within the strategy, rather than baseline confounding. At the same time, they make it impossible to isolate a pure drug effect on delirium independent of the changing medication pathway.

In particular, 4D is not a randomised dexmedetomidine-versus-haloperidol trial. Post hoc division by whether a patient received haloperidol on the inclusion day does not create such a comparison. Haloperidol use depended on agitation and treatment response after allocation. Patients receiving it are consequently selected on clinical features associated with the outcome. The finding that fewer dexmedetomidine patients needed rescue treatment is useful; claims of proven superiority over haloperidol as an alternative first-line regimen go beyond the design.

Blinding and complete primary follow-up among retained participants strengthen confidence in the observed short-term comparison. However, 17 of 168 randomised patients withdrew permission for data use. Their exclusion was legally necessary, but the remaining analysis cannot be assumed identical to an all-randomised treatment comparison. The per-protocol analysis does not recover those missing patients. Sedation and haemodynamic changes could also make treatment allocation partly apparent despite careful preparation procedures, although actual unblinding was not demonstrated.

Stopping followed a stated interim boundary, and the observed P value crossed it. That is stronger than an unplanned decision based on a favourable secondary outcome. Nevertheless, stopping at approximately half the intended sample can leave the estimated benefit exaggerated and rare harms poorly characterised. The primary result warrants attention, but replication would help establish its magnitude. The monitoring account also needs clarification: the principal text says a safety board was not convened because safety outcomes did not differ, while the appendix refers to a board report. This reporting ambiguity does not erase the numerical result, but limits a confident reconstruction of oversight.

The safety data should not be summarised as absence of haemodynamic cost. Treated hypotension was approximately ten percentage points more frequent with dexmedetomidine, and its confidence interval included a clinically important increase. The non-significant comparison cannot establish equal safety. Very few bradycardia events required treatment, but the trial was too small to exclude important differences in uncommon complications. Close monitoring is intrinsic to the tested setting, and the result does not justify extending an infusion strategy to an unmonitored ward.

Applicability is narrower than the term hyperactive delirium might suggest. Median agitation was mild, and patients recently extubated, recently treated with the main study-related drugs, experiencing alcohol withdrawal or unable to undergo reliable assessment were excluded. These exclusions remove several common clinical scenarios. Conversely, recruiting non-intubated patients specifically addresses a setting in which evidence has been sparse. The trial is most relevant when agitation persists despite initial measures and a monitored infusion is being considered to limit further sedative escalation.

Finally, delirium assessment and agitation assessment occurred at different intervals, and the first negative delirium test was not the same as the sustained resolution used to stop treatment. Sedation can alter behaviour without correcting inattention or disorganised thinking. The lack of a clear delirium-duration effect is therefore consistent with the pharmacological distinction between calming a patient and reversing acute brain dysfunction. The proportionate conclusion is that dexmedetomidine can be considered for selected monitored patients needing agitation control, while claims of disease modification, intubation prevention or survival benefit remain unproven.

Body of Evidence

DahLIA provided a direct randomised signal in a different airway situation. Seventy-four ventilated patients whose agitation and delirium prevented extubation were randomised to dexmedetomidine or placebo added to standard care; 71 were analysed. Ventilator-free time over seven days increased by a reported median 17 hours (95% confidence interval 4.0–33.2). Time to extubation and delirium resolution also favoured dexmedetomidine among secondary outcomes. These patients had already been intubated and were selected because agitation obstructed liberation. The result supports a role at that specific transition, but does not establish prevention of intubation in patients breathing without an artificial airway.¹

Carrasco and colleagues studied non-intubated patients more directly, but without random allocation. All 132 initially received haloperidol; 46 non-responders received dexmedetomidine, while 86 responders formed the haloperidol comparison group. Time at satisfactory sedation levels was higher with dexmedetomidine, 92.7% versus 59.3%. Because allocation followed response to the initial drug, the groups differed by construction. The study supported the feasibility of rescue dexmedetomidine and motivated a better-controlled trial, but could not establish comparative first-line efficacy. 4D improves on this evidential limitation through randomisation and masking.²

Larger antipsychotic trials clarify why agitation control and delirium recovery must remain separate claims. MIND-USA randomised 566 patients with respiratory failure or shock and delirium to haloperidol, ziprasidone or placebo. Most had hypoactive delirium. Neither active treatment improved days alive without delirium or coma over 14 days; the corresponding odds ratios were 0.88 and 1.04, with confidence intervals crossing one. Its population was poorly suited to answering rapid control of dangerous agitation specifically, but it challenged routine antipsychotic treatment as a means of shortening delirium.⁵

MENDS2 addressed choice of sedative in ventilated adults with sepsis. Among 422 patients included in the analysis, dexmedetomidine and propofol produced similar adjusted days alive without delirium or coma: 10.7 and 10.8, respectively, with an odds ratio of 0.96 (95% confidence interval 0.74–1.26). Ventilation, survival and later cognitive outcomes did not differ convincingly. This was a light-sedation strategy comparison, not rescue treatment for hyperactive delirium. It argues against assuming a universal cerebral recovery advantage from dexmedetomidine while leaving room for benefit in a specific agitation-management problem.⁶

AID-ICU then randomised 1,000 acutely admitted ICU patients with delirium to haloperidol or placebo. The primary outcome, days alive outside hospital by day 90, did

not improve significantly: the adjusted mean difference was 2.9 days (95% confidence interval –1.2 to 7.0). A lower mortality estimate was a secondary finding and should not be treated as confirmation of the primary hypothesis. This broader trial again illustrates how a drug's effects on symptoms, delirium and later outcomes may differ. It provides context for haloperidol use, rather than a direct head-to-head comparison with 4D's intervention.⁷

Together, these studies support a targeted interpretation of 4D. It fills an important randomised-evidence gap for non-intubated patients, with better early agitation control and less rescue medication under its protocol. It does not overturn the wider uncertainty about pharmacological treatment of delirium itself. The next clinical question is which patients gain enough from that short-term control to justify a monitored infusion and its haemodynamic burden, and whether a simpler or shorter strategy can achieve the same benefit.

Should dexmedetomidine be considered for persistent hyperactive delirium in non-intubated ICU patients?

Consider it for agitation requiring treatment in a closely monitored ICU. 4D supports faster agitation control, but does not establish shorter delirium or fewer intubations.

References

1. Reade MC, Eastwood GM, Bellomo R, Bailey M, Bersten A, Cheung B, et al. Effect of Dexmedetomidine Added to Standard Care on Ventilator-Free Time in Patients With Agitated Delirium: A Randomized Clinical Trial. *JAMA*. 2016;315(14):1460-1468.
2. Carrasco G, Baeza N, Cabré L, Portillo E, Gimeno G, Manzanedo D, et al. Dexmedetomidine for the Treatment of Hyperactive Delirium Refractory to Haloperidol in Nonintubated ICU Patients: A Nonrandomized Controlled Trial. *Crit Care Med*. 2016;44(7):1295-1306.
3. Godet T, Louis C, Rieu B, De Jong A, Couhault P, Pradel G, et al. Dexmedetomidine for treatment of hyperactive delirium in non-intubated ICU patients: the 4D randomized clinical trial. *Intensive Care Med*. 2025;51(12):2305-2317.
4. Godet T, Louis C, Rieu B, De Jong A, Couhault P, Pradel G, et al. Supplementary methods and results to Dexmedetomidine for treatment of hyperactive delirium in non-intubated ICU patients: the 4D randomized clinical trial. *Intensive Care Med*. 2025.

- 
5. Girard TD, Exline MC, Carson SS, Hough CL, Rock P, Gong MN, et al. Haloperidol and Ziprasidone for Treatment of Delirium in Critical Illness. *N Engl J Med*. 2018;379(26):2506-2516.
 6. Hughes CG, Mailloux PT, Devlin JW, Swan JT, Sanders RD, Anzueto A, et al. Dexmedetomidine or Propofol for Sedation in Mechanically Ventilated Adults with Sepsis. *N Engl J Med*. 2021;384(15):1424-1436.
 7. Andersen-Ranberg NC, Poulsen LM, Perner A, Wetterslev J, Estrup S, Hästbacka J, et al. Haloperidol for the Treatment of Delirium in ICU Patients. *N Engl J Med*. 2022;387(26):2425-2435.

DISTAL

Psychogios M, Brehm A, Ribo M, Rizzo F, Strbian D, Rätty S, et al. Endovascular Treatment for Stroke Due to Occlusion of Medium or Distal Vessels. *N Engl J Med*. 2025;392(14):1374-1384.

Introduction

Successful treatment of a proximal arterial occlusion does not establish that the same procedure benefits every smaller branch occlusion. Moving a thrombectomy device into a narrower, more tortuous vessel changes the technical challenge. The amount and functional importance of threatened brain also vary, while the risks of vessel injury and intracranial bleeding remain. The relevant decision is whether adding an invasive procedure improves the patient's eventual recovery enough to justify those risks.

The large-vessel thrombectomy trials created a strong incentive to extend treatment. They also created a risk of assuming that anatomical proximity implies therapeutic equivalence. A branch labelled M2 can supply a substantial territory, while another branch with the same segment label may be smaller or nondominant. A severe language deficit may be disabling despite a modest aggregate stroke score. Conversely, a visible occlusion can coexist with a favourable prospect of recovery after medical treatment. Neither the angiogram nor the clinical score alone supplies the complete treatment decision.

Earlier evidence concerning M2 occlusions was encouraging but came from a small, selected subgroup of trials designed principally around more proximal disease. It did not provide a secure answer for the full range of medium and distal vessels. Dedicated randomised comparisons were therefore necessary before an invasive treatment could be justified routinely in these patients. DISTAL contributes to that decision alongside the contemporaneous ESCAPE-MeVO trial.^{1,2,3}

Synopsis

DISTAL randomised 553 patients at 55 hospitals; 543 remained after consent exclusions, with 271 allocated to thrombectomy plus medical treatment and 272 to medical treatment. Eligible occlusions included nondominant or codominant M2 and more distal anterior or posterior circulation vessels. Treatment was permitted within 24 hours, with salvageable tissue required beyond six hours.⁴

Median age was 77 years and stroke severity score six; 65.4% received intravenous thrombolysis. Assessment was masked. The primary 90-day modified Rankin scale analysis found no favourable disability shift: common odds ratio 0.90 (95% confidence interval 0.67–1.22), $P=0.50$. Mortality was 15.5% versus 14.0%; symptomatic intracranial haemorrhage occurred in 5.9% versus 2.6%. Thrombectomy was performed in 84.5% of those assigned to it, with successful reperfusion reported in 71.7%. Device choice was discretionary. Secondary outcomes did not establish benefit. The primary analysis used imputation and combined severe dependency and death; secondary testing followed a gatekeeping hierarchy.⁴

Critique

The clinical conclusion concerns an offer of thrombectomy within the tested selection and delivery strategy. It does not establish that restoring flow to an individual threatened brain region is biologically useless. Those are different questions. Randomisation evaluates the net consequences of identifying a patient, arranging the procedure, attempting treatment and managing its complications. A technically attractive intervention must succeed at that level to justify routine use.

The primary estimate provides no affirmative basis for expanding treatment. Its uncertainty leaves room for a smaller benefit as well as harm, so a negative superiority result should not be called equivalence. Equally, the possibility of benefit somewhere within a confidence interval is not a positive recommendation. For an invasive procedure, the burden of evidence rests on demonstrating a favourable balance of recovery and complications, not on requiring a trial to exclude every conceivable advantage before practice is restrained.

Anatomical classification is central to applying the result. Dominance describes the territory supplied, whereas a segment label describes position. These concepts should not be substituted for one another when translating evidence into a treatment pathway. A policy that treats all M2 occlusions identically risks combining substantially different clinical problems. Conversely, a policy that defines every branch lesion as too small to matter can ignore an eloquent territory and a disabling deficit. The appropriate response is more precise selection, rather than an indiscriminate expansion or contraction of established proximal-vessel care.

A global stroke severity score is useful for describing a population but cannot fully represent individual disability. Language, vision, hand function and gait carry different consequences for different patients. That observation justifies careful clinical assessment; it does not validate an untested subgroup recommendation. To

demonstrate that a particular deficit identifies treatment benefit requires an appropriately designed comparison, not simply a compelling account of how disabling that deficit can be.

The control strategy is another important consideration. Best medical treatment is an active alternative that can include thrombolysis, physiological management, stroke-unit care and rehabilitation. A comparison against that pathway asks whether catheter treatment adds enough value to justify its use. Observational series of technically successful procedures often lack a comparable untreated group and therefore cannot establish what would have happened without intervention. Good outcomes after a procedure are insufficient evidence of procedural benefit when recovery can also occur through the competing pathway.

Treatment receipt should be interpreted within this strategy question. Some patients assigned to an intervention may improve, recanalise or become unsuitable before the procedure is attempted. Excluding them from the efficacy analysis can destroy the protection offered by randomisation, because the reason for not proceeding is associated with prognosis. An analysis restricted to successfully treated patients would be even more selective. It could describe technical performance, but would not answer whether assigning the strategy improves outcome.

The relationship between reperfusion and functional recovery also requires care. Angiographic opening is an intermediate outcome. Brain tissue may already be irreversibly injured, reperfusion may be incomplete at the microvascular level, or procedural harm may offset a physiological gain. Demonstrating a better technical result with a future device would therefore be encouraging, but would still require confirmation that patients function better. A new device should not inherit a clinical efficacy claim solely because it is smaller or easier to navigate.

Timing creates a related challenge. There is a plausible reason to pursue faster diagnosis and treatment in any reperfusion strategy, but an explanation based on delay needs direct evidence. Comparing faster and slower treated patients can be confounded by anatomy, workflow, clinical stability and ease of access. A negative strategy trial cannot be converted into a positive one by assuming that an idealised service would have achieved a different result. Faster delivery is a testable improvement, not a substitute for demonstrated efficacy.

The disability endpoint is appropriate to the purpose of stroke treatment, while still leaving some questions unresolved. An ordinal scale captures more of the distribution than a single favourable-outcome threshold, but assumes that the model meaningfully

summarises change across categories. It does not measure every cognitive, visual or occupational consequence. Additional patient-centred measures can enrich interpretation, yet choosing whichever secondary measure appears most favourable after a negative primary result risks selective inference. The outcome hierarchy should remain intact.

Safety comparisons need the same discipline. A numerically greater bleeding rate should not be dismissed because an individual safety test lacks statistical significance. Small numbers produce wide intervals, particularly for uncommon but serious complications. At the same time, a numerical difference alone cannot establish the size or mechanism of harm. The appropriate judgement considers the entire benefit–harm pattern and avoids describing the procedure as equally safe when the trial was not designed to establish safety equivalence.

Subgroup exploration can help formulate the next question but cannot provide a reliable escape from an unfavourable overall result. Vessel location, symptom severity, imaging appearance and thrombolysis eligibility offer many possible divisions. With enough comparisons, some point estimates will look promising by chance. A credible claim needs an interaction, biological coherence, sufficient observations and preferably independent confirmation. A within-group confidence interval or an attractive forest-plot point is not enough to identify a clinical responder population.

For critical care practice, the decision also extends beyond the angiography suite. Sedation or anaesthesia, haemodynamic management, transport and subsequent monitoring form part of the intervention pathway. Their benefits and burdens should be included when comparing strategies. This does not imply that procedural support caused the observed result; it means that a treatment recommendation must concern the complete real-world process rather than the device in isolation.

The appropriate practical response is to preserve established indications while resisting unsupported extension. A neutral trial in a newly studied vascular territory does not negate the large-vessel evidence. Nor does the success of large-vessel thrombectomy override a dedicated randomised comparison in a different population. Where an exceptional case prompts consideration of treatment, the uncertainty should be recognised explicitly, with a preference for research capable of resolving the selection question.

Body of Evidence

The original HERMES individual-patient meta-analysis pooled five randomised trials and 1,287 patients with predominantly proximal anterior-circulation occlusions. Thrombectomy improved the 90-day disability distribution, with an adjusted common odds ratio of 2.49 (95% confidence interval 1.76–3.53). This established a major treatment effect within the populations studied. It supplied the rationale for further investigation, but did not establish that the same benefit persisted as vessel calibre and threatened territory decreased. Its participants should not be counted again as independent evidence when subsequent subgroup analyses are discussed.¹

The HERMES M2 analysis examined 130 patients, most with proximal M2 occlusions. Functional independence occurred in 58.2% after thrombectomy and 39.7% with control treatment; the adjusted odds ratio was 2.39 (95% confidence interval 1.08–5.28). The subgroup was small and drawn from the parent randomised trials, with the largest apparent effects in proximal or dominant branches. It therefore supported treatment consideration for selected substantial M2 territories more clearly than a general policy for every medium or distal branch. The distinction between shared parent data and a new dedicated trial is crucial when weighing the strength of this evidence.²

ESCAPE-MeVO provides an independent, contemporary test of a related clinical question. It randomised 530 patients across five countries, using a treatment window within 12 hours and favourable baseline imaging. Its primary outcome was an excellent functional result, defined as a modified Rankin scale score of 0 or 1 at 90 days. This occurred in 41.6% assigned to thrombectomy and 43.1% assigned to usual care, with an adjusted rate ratio of 0.95 (95% confidence interval 0.79–1.15). Thus, the result did not support routine intervention in that selected medium-vessel population.³

The safety findings in ESCAPE-MeVO reinforce the need for a net clinical benefit. Mortality was 13.3% with thrombectomy and 8.4% with usual care, with an adjusted hazard ratio of 1.82 (95% confidence interval 1.06–3.12). Symptomatic intracranial haemorrhage was also numerically more frequent. These outcomes require interpretation within their secondary status and the trial's full analysis, but they prevent a reassuring conclusion based solely on the absence of primary efficacy. Lack of benefit and uncertainty about harm are not equivalent to a harmless procedure awaiting a better efficacy endpoint.³

The two dedicated trials should be compared through their clinical questions rather than by placing their effect measures side by side as if they were identical. A common odds ratio for an ordinal disability distribution and a rate ratio for an excellent-outcome threshold answer related but distinct questions. Different vessel definitions, imaging

requirements and treatment windows also affect the enrolled populations. Agreement in the broad clinical conclusion is informative even when the numerical estimates cannot be directly equated.

At the end of 2025, the evidence supports caution about extending a successful proximal-vessel treatment to smaller occlusions as a routine policy. The remaining research need is specific: identify a population and an intervention pathway in which the added recovery benefit outweighs procedural harm. That may involve better anatomical classification, selection based on threatened tissue, faster delivery or different technology. Each is a hypothesis to test prospectively, rather than an established reason to disregard the dedicated randomised evidence.

Should thrombectomy routinely be added to medical treatment for medium or distal vessel occlusion?

No. The dedicated trials do not establish improved functional recovery, and a clinically important risk of harm remains.

References

1. Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet*. 2016;387(10029):1723-1731.
2. Menon BK, Hill MD, Davalos A, Roos YBWEM, Campbell BCV, Dippel DWJ, et al. Efficacy of endovascular thrombectomy in patients with M2 segment middle cerebral artery occlusions: meta-analysis of data from the HERMES Collaboration. *J Neurointerv Surg*. 2019;11(11):1065-1069.
3. Goyal M, Ospel JM, Ganesh A, Dowlatshahi D, Volders D, Möhlenbruch MA, et al. Endovascular Treatment of Stroke Due to Medium-Vessel Occlusion. *N Engl J Med*. 2025;392(14):1385-1395.
4. Psychogios M, Brehm A, Ribo M, Rizzo F, Strbian D, Rätty S, et al. Endovascular Treatment for Stroke Due to Occlusion of Medium or Distal Vessels. *N Engl J Med*. 2025;392(14):1374-1384.

BRIDGE-TNK

Qiu Z, Li F, Sang H, Yuan G, Xie D, Zhou K, et al. Intravenous Tenecteplase before Thrombectomy in Stroke. *N Engl J Med*. 2025;393(2):139-150.

Introduction

Rapid thrombectomy does not make intravenous thrombolysis automatically redundant. A thrombolytic may restore flow before catheter treatment, facilitate clot removal or act on smaller downstream thrombi that remain after the main vessel is opened. It can also cause bleeding, and its administration must fit into a time-critical pathway. For an otherwise eligible patient arriving directly at a thrombectomy centre, the practical question is whether omitting the intravenous treatment sacrifices meaningful recovery.

Trials of direct thrombectomy versus bridging thrombolysis produced a more uncertain answer than a simple comparison of non-significant results suggested. Individual non-inferiority trials used different margins and did not all meet their criteria. A patient-level synthesis failed to establish that direct thrombectomy preserved an acceptably similar outcome. Most of this evidence concerned alteplase. It could not definitively determine the incremental value of tenecteplase, which can be administered as a single bolus.¹

EXTEND-IA TNK had compared tenecteplase with alteplase before planned thrombectomy and found more early reperfusion with tenecteplase. That is a drug-selection question, distinct from whether thrombolysis should be given at all. BRIDGE-TNK directly tested the latter question for tenecteplase: does giving 0.25 mg/kg before prompt thrombectomy improve functional independence compared with proceeding to thrombectomy alone?^{2,3}

Synopsis

BRIDGE-TNK was an investigator-initiated, open-label, randomised superiority trial with masked outcome assessment at 39 hospitals in China. Recruitment ran from May 2022 to September 2024. Public and institutional funding supported the study, and CSPC Recomgen supplied the tenecteplase formulation and provided support. The manufacturer had no role in design, data collection, analysis or publication decisions. Patients or legal representatives gave written consent before randomisation.^{3,4}

Adults with an acute ischaemic stroke caused by an eligible large-vessel occlusion were enrolled within 4.5 hours of last known well if they were eligible for intravenous thrombolysis and thrombectomy was planned. The principal report describes internal carotid, M1, M2 and vertebrobasilar occlusions; the supplementary criteria also allowed P1 occlusion. Substantial pre-stroke disability was excluded: the permitted modified Rankin scale score was at most two below age 80 and at most one in older patients. Other exclusions included contraindications to thrombolysis, prior thrombolysis for the presenting stroke, unsuitable arterial access, occlusions in multiple vascular territories and conditions preventing reliable outcome assessment.

Patients were assigned equally through a secure website using fixed blocks of four. The patients and clinical teams knew allocation, but independent neurological assessors, the imaging core laboratory and the clinical-events committee were masked. The trial concerned patients presenting directly to thrombectomy-capable hospitals; it did not test an interhospital transfer strategy after thrombolysis.

Tenecteplase was administered intravenously at 0.25 mg/kg, to a maximum of 25 mg, over five to ten seconds followed by a saline flush. Both groups were to receive rapid endovascular treatment. The choice of stent retriever, aspiration, angioplasty, stenting and intra-arterial adjuncts was left to the interventionalist. The control group proceeded without intravenous tenecteplase. Thus, the intervention was a bolus integrated into an urgent procedural pathway, with no planned waiting period to assess whether thrombolysis worked.

The primary outcome was functional independence at 90 days, defined as a modified Rankin scale score of 0–2. Two certified independent neurologists adjudicated structured video or voice assessments while masked to allocation. The scale ranges from no symptoms to death; a score of two allows independence in daily affairs despite some residual disability. Secondary outcomes included the full disability distribution, excellent outcome at a score of 0–1, independent walking at 0–3, quality of life, neurological deficit and angiographic reperfusion before and after thrombectomy. Safety outcomes included symptomatic intracranial haemorrhage within 48 hours, mortality and procedural complications.

Planning assumed functional independence in 41% after direct thrombectomy and 54% after bridging tenecteplase, an absolute improvement of 13 percentage points. A sample of 462 would provide 80% power at a two-sided 5% significance level; the target increased to 544 to allow 15% attrition. Modified Poisson regression estimated risk ratios adjusted for age, baseline stroke severity, baseline Alberta Stroke Program Early Computed Tomography Score, occlusion site and time to randomisation. Both adjusted

and unadjusted results were presented. The ordinal disability analysis used a generalised odds ratio, and skewed continuous outcomes used win ratios. Secondary and subgroup intervals were not adjusted for multiplicity.

A total of 554 patients were randomised. Four withdrew consent, leaving 550 for analysis: 278 assigned to tenecteplase plus thrombectomy and 272 to thrombectomy alone. No retained patient was lost to 90-day follow-up, and primary outcomes and adjustment covariates were complete. Median age was 70 years, 41.8% were women, and median National Institutes of Health Stroke Scale score was 16. Median baseline imaging score was eight. Approximately 95% had no pre-stroke symptoms or disability. Cardioembolism accounted for about 61% of strokes and large-artery atherosclerosis for about 30%.

Tenecteplase started a median of six minutes after allocation. The median interval from the bolus to arterial puncture was 16 minutes. Randomisation-to-puncture times were 28.2 minutes with bridging and 24.4 minutes with direct thrombectomy; puncture-to-reperfusion times were 55 and 64 minutes. Angioplasty or stenting was used in 51 of 261 patients with procedural data in the bridging group and 72 of 262 controls. The smaller procedure-time values in the supplement exclude patients receiving angioplasty or stenting and are not conflicting estimates for the whole cohort.^{3,4}

Functional independence occurred in 147 of 278 bridging patients (52.9%) and 120 of 272 controls (44.1%), an observed difference of 8.8 percentage points. The unadjusted risk ratio was 1.20 (95% confidence interval 1.01–1.43); the prespecified adjusted risk ratio was 1.18 (95% confidence interval 1.01–1.39), with $P=0.04$ for both. The full disability distribution was less conclusive, with an adjusted generalised odds ratio of 1.16 (95% confidence interval 0.94–1.43). Excellent outcome occurred in 34.9% versus 27.9%, while independent walking occurred in 64.0% versus 61.8%. Quality of life did not favour bridging on its win-ratio analysis.³

Successful reperfusion before thrombectomy occurred in 17 of 278 bridging patients (6.1%) and three of 271 controls (1.1%); the adjusted risk ratio was 5.19 (95% confidence interval 1.51–17.84). Final successful reperfusion was common in both groups, 91.4% and 94.1%, with an adjusted risk ratio of 0.97 (95% confidence interval 0.92–1.02). First-pass outcomes were also similar. The early angiographic difference therefore did not translate into a higher final macrovascular reperfusion rate.

Symptomatic intracranial haemorrhage occurred in 23 of 271 patients with available data after bridging (8.5%) and 18 of 269 controls (6.7%); the adjusted risk ratio was 1.35 (95% confidence interval 0.74–2.44). Ten patients lacked central haemorrhage data, although

local radiologists had not reported haemorrhage. Death by day 90 occurred in 62 of 278 (22.3%) and 54 of 272 (19.9%), with an adjusted hazard ratio of 1.17 (95% confidence interval 0.81–1.69). Other complications and serious events did not show a clear pattern of benefit. Subgroup analyses were exploratory, including comparisons based on actual tenecteplase timing in one arm and estimated timing in the other.

Critique

BRIDGE-TNK provides direct evidence of improved functional independence with tenecteplase before thrombectomy in its early-window population. The endpoint was prespecified, clinically important and assessed by masked adjudicators, so it should not be dismissed merely because the P value was close to 0.05. Nevertheless, the confidence interval lies close to no benefit at its lower boundary, and other functional measures do not supply consistent confirmation. The appropriate interpretation is supportive evidence with material uncertainty about the size and breadth of benefit.³

The choice of a binary independence threshold matters. Crossing from dependence to independence is valuable, but a favourable result at that boundary need not mean improvement throughout the disability distribution. The comparison at a threshold permitting independent walking was much smaller, the ordinal estimate included no effect, and mortality was numerically higher with bridging. These observations do not invalidate the primary outcome. They do make a broad claim of uniformly better neurological recovery less secure than the primary headline alone suggests.

The trial was designed to detect a larger absolute improvement than it observed. Its complete follow-up among retained patients is a substantial strength, particularly given the originally generous attrition allowance. Adjustment and crude analysis agreed, reducing concern that a selected model created the main result. Four consent withdrawals are unlikely to dominate interpretation, although the population still excludes some randomised patients. There is no reason to label the trial simply underpowered after it met its primary significance criterion; the issue is the precision and consistency of the estimated benefit.

The proposed mechanisms remain incompletely resolved. More early reperfusion is consistent with a pharmacological effect before catheter treatment. A shorter procedure could also plausibly preserve tissue, and lysis of distal emboli might improve microvascular flow without changing a broad angiographic grade. However, the trial did not randomise those intermediate events. Differences in procedure duration, rescue treatment and early recanalisation cannot establish which pathway caused the functional

result. The modest absolute frequency of pre-procedure reperfusion also means it should not be assumed to explain the entire independence difference.

The final reperfusion findings are particularly informative. Both groups achieved high rates of successful macrovascular reperfusion, so this was a test of incremental benefit in a setting where thrombectomy was generally effective. A clinician should not infer that thrombolysis is unnecessary once a high procedural success rate is expected. Equally, the positive independence result does not prove that tenecteplase improves the technical endpoint measured after thrombectomy. Functional recovery and vessel opening are related but distinct outcomes.

Open treatment assignment could influence procedural choices and subsequent care. Masking the imaging laboratory and clinical adjudicators reduces assessment bias, and the complete outcome data improve confidence. It does not prevent differences in decisions about angioplasty, stenting or adjunctive medication. These may be consequences of the assigned strategy, genuine chance imbalances or responses to unmeasured procedural features. Their presence calls for careful interpretation rather than an automatic conclusion that the trial is either invalid or mechanistically explained.

The short interval between bolus and puncture is important for applicability. It demonstrates that tenecteplase can be incorporated into a rapid thrombectomy pathway. The trial did not test delaying the procedure to allow the drug more time to work. Comparisons with earlier studies that allowed longer intervals cannot establish an optimal waiting time, because patients, clot location, workflow and comparator drug also differed. A positive bridging result should support prompt parallel treatment, not an intentional pause before thrombectomy.

Safety remains uncertain in clinically relevant directions. The haemorrhage interval allows both a reduction and more than a doubling of symptomatic bleeding risk, while the mortality interval includes a meaningful increase. Neither outcome supports a declaration of equal safety. The missing central haemorrhage assessments are few but relevant because they concern an uncommon serious event; a local report of no bleeding is not identical to complete central adjudication. The primary functional endpoint must therefore be considered alongside an unresolved haemorrhagic burden.

Subgroup findings should not dictate a new timing or aetiological rule. Several strata were examined, some with relatively few events, and the analysis did not control multiplicity. Comparing actual treatment timing in the bridging arm with a constructed expected time in controls adds another layer of uncertainty. A more favourable point estimate beyond a selected time threshold does not mean that waiting improves the

drug's effect. Similarly, a favourable cardioembolic subgroup cannot establish that treatment should be withheld in atherosclerotic stroke without convincing interaction evidence.

The population also limits generalisation. Patients arrived directly at thrombectomy hospitals and were eligible for intravenous thrombolysis within 4.5 hours. The study does not resolve bridging during interhospital transfer, late-window treatment or patients with important contraindications. Posterior-circulation occlusions were a minority. Large-artery atherosclerosis was common, but that category should not be equated with intracranial atherosclerosis in every participant: the supplement distinguishes intracranial and extracranial lesions. Different case mix and procedural practice may alter the incremental benefit elsewhere.

Finally, BRIDGE-TNK compares tenecteplase with omission of intravenous thrombolysis. It does not establish superiority over alteplase, and it should not be presented as a head-to-head choice between thrombolytic agents. The tested formulation, dose and delivery pathway are explicit. Within that scope, the result supports retaining bridging treatment rather than routinely omitting it solely because thrombectomy is imminent, while leaving the magnitude of overall benefit and risk to be clarified by further direct evidence.

Body of Evidence

EXTEND-IA TNK addressed the choice of thrombolytic before thrombectomy. In 202 patients treated within 4.5 hours, tenecteplase 0.25 mg/kg was compared with alteplase 0.9 mg/kg. Reperfusion of more than half the affected territory, or absence of a retrievable thrombus at initial angiography, occurred in 22% versus 10%, an absolute difference of 12 percentage points (95% confidence interval 2–21). The secondary disability analysis also favoured tenecteplase. This supported its use as a bridging agent, but both groups received a thrombolytic, so the trial could not determine whether either drug added value over direct thrombectomy.²

The IRIS collaboration subsequently pooled individual data from six direct-versus-bridging trials involving 2,313 patients, predominantly treated with alteplase when thrombolysis was used. The adjusted common odds ratio for better disability outcome with direct thrombectomy was 0.89 (95% confidence interval 0.76–1.04). The lower bound crossed the prespecified non-inferiority boundary of 0.82, so non-inferiority of omitting thrombolysis was not established. The interval also crossed one, preventing a conventional superiority conclusion. This distinction explains why the

earlier evidence could be described as inconclusive without supporting routine abandonment of bridging treatment.¹

BRIDGE-TNK fills a different part of the evidence map. It gives tenecteplase a direct comparison against no intravenous pretreatment, rather than relying on the combination of alteplase-based omission trials and a separate drug-comparison trial. That is a useful advance: indirect reasoning through two different comparators can be misleading when timing, case mix and treatment effects interact. Its positive primary result strengthens the argument for preserving early bridging treatment, while the less conclusive secondary pattern prevents a claim that all previous uncertainty has disappeared.³

The early reperfusion proportions from different trials should not be ranked as if they measure the intrinsic efficacy of a drug under identical conditions. An angiogram obtained soon after injection gives the drug less time to act than one obtained later. Clot burden, location and prior imaging selection can also differ. A lower early reperfusion proportion in one tenecteplase trial therefore does not demonstrate inferior pharmacological activity, just as a longer interval does not establish a reason to delay a procedure. The patient-centred randomised comparison remains the decisive measure.

The same restraint applies when linking this evidence to other reperfusion strategies. Intravenous treatment before an occluded vessel is opened, intravenous treatment when thrombectomy is not planned, and intra-arterial treatment after successful thrombectomy address different residual problems. A shared fibrinolytic mechanism does not make their indications interchangeable. Evidence supporting one position in the pathway should be carried across to another only when the population, timing, route and comparator justify it.

At the end of 2025, BRIDGE-TNK provides a reason to continue tenecteplase before prompt thrombectomy in appropriately selected early presenters. It supplies neither proof of a mortality benefit nor reassurance that clinically important bleeding differences have been excluded. Further direct comparisons and pooled analyses can improve precision, but their interpretation should retain the distinctions between superiority, non-inferiority, drug selection and treatment omission that shaped the original question.

Should eligible early-presenting patients receive tenecteplase before prompt thrombectomy?

BRIDGE-TNK supports this approach without delaying thrombectomy. The improvement in functional independence is encouraging, although the broader benefit–harm balance remains imprecise.

References

1. Majoie CB, Cavalcante F, Gralla J, Yang P, Kaesmacher J, Treurniet KM, et al. Value of intravenous thrombolysis in endovascular treatment for large-vessel anterior circulation stroke: individual participant data meta-analysis of six randomised trials. *Lancet*. 2023;402(10406):965-974.
2. Campbell BCV, Mitchell PJ, Churilov L, Yassi N, Kleinig TJ, Dowling RJ, et al. Tenecteplase versus Alteplase before Thrombectomy for Ischemic Stroke. *N Engl J Med*. 2018;378(17):1573-1582.
3. Qiu Z, Li F, Sang H, Yuan G, Xie D, Zhou K, et al. Intravenous Tenecteplase before Thrombectomy in Stroke. *N Engl J Med*. 2025;393(2):139-150.
4. Qiu Z, Li F, Sang H, Yuan G, Xie D, Zhou K, et al. Supplementary methods and results to Intravenous Tenecteplase before Thrombectomy in Stroke. *N Engl J Med*. 2025.

ESCAPE-MeVO

Goyal M, Ospel JM, Ganesh A, Dowlathshahi D, Volders D, Möhlenbruch MA, et al. Endovascular Treatment of Stroke Due to Medium-Vessel Occlusion. N Engl J Med. 2025;392(14):1385-1395.

Introduction

The success of thrombectomy for proximal large-vessel stroke creates a powerful temptation to extend treatment further into the cerebral circulation. An occluded smaller artery still deprives brain tissue of blood, and a relatively small infarct can cause disabling aphasia, visual loss or loss of hand function. The anatomical size of the vessel does not reliably describe the importance of the deficit to the individual patient.

However, the balance of benefit and harm may change as the target becomes smaller and more distal. Less tissue may be available to save, intravenous thrombolysis may be more effective and spontaneous improvement may be more common. At the same time, navigating and withdrawing devices through smaller arteries can remain technically demanding. A treatment that produces a large net benefit in proximal occlusion can therefore have little benefit, or cause net harm, in another part of the vascular tree.

Observational experience cannot reliably settle this question. Operators select patients according to anatomy, deficit, time, prognosis and their own expectation of success. Comparisons between treated and untreated patients may retain substantial confounding even after statistical adjustment. Dedicated randomised trials are needed to establish whether the invasive strategy improves recovery beyond contemporary medical care.

ESCAPE-MeVO directly tested that extension of thrombectomy. Together with DISTAL, it is a warning against allowing procedural feasibility and success in an adjacent population to stand in for evidence of patient benefit. It also leaves a substantial unmet need: declining to perform an unproven procedure does not make the consequences of medium-vessel stroke trivial.^{1,2}

Synopsis

ESCAPE-MeVO was a phase 3, multicentre, randomised, open-label trial with blinded outcome assessment. Between April 2022 and June 2024, 530 patients were enrolled at 58 sites in Canada, the United States, Germany, the United Kingdom and Hungary. Funding came from the Canadian Institutes of Health Research and Medtronic. The investigators reported that funders were not involved in design, conduct, data analysis or manuscript preparation. An independent committee monitored safety and efficacy.^{1,3}

Eligible adults had an acute stroke caused by an M2 or M3 middle cerebral artery occlusion, an A2 or A3 anterior cerebral artery occlusion, or a P2 or P3 posterior cerebral artery occlusion. A1 and P1 occlusions were excluded. Randomisation had to occur within 12 hours of last known well. NIHSS was greater than 5, or 3–5 with a deficit considered disabling in the context of the patient's occupation and circumstances. The neurological deficit had to correspond to the vascular lesion.¹

Baseline imaging had to support the presence of salvageable tissue. Assessment followed local workflows using CT or MRI, vascular imaging and, where available, perfusion imaging. Selection incorporated the extent of established ischaemic change, collateral supply or visual evidence of a core–penumbra mismatch. The trial therefore allowed several imaging routes to eligibility rather than requiring a single automated perfusion threshold in every patient. Patients requiring daily nursing care or assistance with activities of daily living before stroke were excluded.^{1,2}

A central web-based dynamic allocation algorithm balanced age, sex, baseline NIHSS, vascular territory and site. Patients were assigned to thrombectomy plus usual care or usual care alone. Both groups received intravenous thrombolysis when indicated, stroke-unit care, rehabilitation and secondary prevention. The thrombectomy strategy required a Solitaire X family stent retriever for the first retrieval attempt, with or without aspiration. Subsequent devices and techniques were discretionary. General anaesthesia was encouraged to facilitate safe distal navigation.¹

The trial was planned around the 90-day mRS distribution, with scores 5 and 6 combined. The report states that a prespecified fallback would promote mRS 0–1 to the primary outcome if the proportional-odds model was unsuitable. That assumption was not met, and the main reported comparison was therefore the proportion achieving an excellent outcome. Masked trained assessors measured the mRS. Secondary endpoints included mRS 0–2, mortality, independence in daily activities, quality of life and follow-up infarct volume.¹

The intended sample was 500, increased to 530 to allow for attrition. An interim analysis after 250 patients had completed follow-up recommended continuation. Alpha spending left a final efficacy threshold of $P < 0.0398$. Adjusted binary outcomes were estimated using a Poisson model with robust variance, controlling for the balancing variables. Mortality used a Cox model. Secondary and subgroup analyses were exploratory and were not adjusted for multiplicity. Efficacy was principally analysed by randomised group, while the reported safety comparison used treatment received.¹

Of 530 randomised patients, 255 were assigned to thrombectomy and 275 to usual care. One control patient immediately withdrew and was omitted from the efficacy population. Three controls crossed to thrombectomy and one intervention patient crossed to usual care. Eight primary outcomes were imputed: four from the 30-day mRS and four as mRS 4 when patients were known to be alive but both functional assessments were unavailable. Protocol deviations affected 52 participants; the per-protocol population contained 477.^{1,3}

Median age was 75 years and median baseline NIHSS was 8 overall. Approximately 85% of occlusions involved a middle cerebral artery branch. Both proximal and distal M2 lesions were represented, as were M3 lesions. Intravenous thrombolysis was given to 56.5% of the intervention group and 60.2% of controls. Median onset-to-randomisation time was approximately 4.5 and 4.2 hours, respectively.¹

Successful reperfusion was reported in 190/253 intervention patients, 75.1%, using a medium-vessel adaptation of the eTICI scale. Thirty-nine intervention patients had already recanalised by the first angiogram and did not require clot retrieval; 32 of these had received intravenous thrombolysis. Median hospital-arrival-to-arterial-access time was 95 minutes, and onset-to-final angiography or reperfusion was approximately six hours. The supplementary procedural table reports general anaesthesia in 105 patients, 41.3% of those with sedation data.^{1,3}

An excellent outcome at 90 days occurred in 106/255 intervention patients and 118/274 controls, 41.6% versus 43.1%. The adjusted rate ratio was 0.95 (95% CI 0.79–1.15; $P = 0.61$). Functional independence, mRS 0–2, occurred in 54.1% versus 58.8%, adjusted rate ratio 0.92 (95% CI 0.80–1.05). Neither comparison established benefit from the invasive strategy.¹

Mortality was 34/255 versus 23/274, 13.3% versus 8.4%. The adjusted hazard ratio was 1.82 (95% CI 1.06–3.12), whereas the unadjusted estimate was 1.62 (95% CI 0.95–2.75). A Barthel Index of at least 95 occurred in 53.5% versus 64.7%, also favouring usual care,

although this was an exploratory endpoint with missing assessments. There was no clear benefit in quality of life or infarct volume.¹

The safety population comprised 257 patients treated with the endovascular strategy and 272 receiving usual care. Symptomatic intracranial haemorrhage occurred in 14/257 versus 6/272, 5.4% versus 2.2%. Serious adverse events affected 33.9% versus 25.7%. Pneumonia, recurrent stroke, stroke progression and symptomatic haemorrhage were among the common events. These safety denominators differ from the randomised efficacy denominators and should be retained when quoting the results.¹

Critique

ESCAPE-MeVO addresses the relevant clinical decision: whether to add an invasive treatment pathway to contemporary medical care. Central allocation, several participating countries, masked clinical assessment and low crossover strengthen that comparison. The trial reached its target after independent interim review. Its result cannot be dismissed simply because thrombectomy was technically possible or because previous uncontrolled experience suggested benefit.

The primary confidence interval is compatible with modest benefit or harm, so the trial does not prove exact equivalence. It does, however, fail to show the large benefit that might have been anticipated from proximal large-vessel experience. The point estimate favours usual care, and the independence, daily-activity and mortality results provide no convincing countervailing signal of patient benefit. Routine intervention requires affirmative evidence of net benefit; uncertainty around a null result does not supply that evidence.

The fallback from an ordinal analysis to mRS 0–1 needs transparent explanation. The publication describes the decision rule as prespecified, triggered by failure of the proportional-odds assumption. That is different from choosing an endpoint solely because its P value is more favourable. Nevertheless, a binary excellent-outcome endpoint answers a narrower question than improvement across the entire disability distribution, and the sample-size plan was based on the ordinal outcome. The change in analysis should remain visible rather than presenting the binary endpoint as the only original design consideration.

The result also reflects a treatment strategy rather than an isolated device manoeuvre. Some patients had already recanalised when angiography was performed. They could not gain from retrieving a clot that was no longer present, yet still experienced the transfer, preparation and angiographic pathway. Including them in the randomised

analysis is appropriate because clinicians must decide whether to initiate that pathway before knowing the final angiographic finding. Removing them after allocation would answer a different and potentially biased question.

Technical success was achieved in about three-quarters of intervention patients. That leaves scope for better devices or technique, but it also describes the practical effectiveness of the approach tested. An analysis restricted to patients with successful reperfusion would select on a post-randomisation outcome associated with anatomy, operator performance and prognosis. It could not establish what would happen if all eligible patients were offered thrombectomy. A better technical approach needs its own evaluation rather than a favourable reinterpretation of selected procedural successes.

The first-pass device requirement limits extrapolation to other strategies. Aspiration-first treatment, different stent retrievers or future devices may produce different technical and safety results. The trial does not prove that they cannot work. It equally provides no evidence that they do work. Device development should be accompanied by randomised patient-outcome testing, because a more convenient or successful angiographic procedure need not deliver sufficient clinical benefit to offset harm.

The mortality difference deserves attention without being overstated. It was a secondary outcome with no multiplicity adjustment, and its conventional statistical significance depended on covariate adjustment. The trial was not designed primarily to detect a mortality difference. Nonetheless, a higher death rate alongside more symptomatic haemorrhage and serious adverse events is clinically concerning, especially in a trial with no demonstrated functional advantage. Calling the study merely neutral would obscure that concern.

The data do not establish that general anaesthesia caused the adverse outcome pattern. Anaesthetic choice was not randomised and may have reflected technical difficulty, agitation, illness severity or local practice. Pneumonia was more frequent in the intervention group, but attributing this to one procedural component would require stronger evidence. The appropriate inference is that the whole invasive strategy carried additional burdens and possible harms; the exact contribution of each component remains uncertain.

Workflow is another plausible modifier rather than a proven explanation. Reperfusion took approximately six hours from onset, and the authors contrast this with faster pathways in earlier large-vessel trials. Smaller-vessel stroke may involve more nuanced decisions, more frequent initial improvement and technically difficult access. Faster

treatment could conceivably change the balance. It does not follow that this trial would have been positive had a particular interval been shortened, because that counterfactual was not randomised.

The exploratory time subgroup should be interpreted in the same way. The earlier group had a point estimate favouring intervention and the later group one favouring control, but the early-group interval included no effect. Timing was divided at the observed median, and several subgroup comparisons were made. These findings support studying more rapid treatment; they do not establish a validated time threshold below which routine intervention is beneficial.³

Applicability is strongest to the enrolled population, predominantly older patients with middle cerebral artery branch occlusions and moderate deficits. There were too few anterior cerebral and some posterior cerebral artery lesions for reliable vessel-specific estimates. No screening logs were required, so the investigators could not determine how many otherwise relevant patients were excluded or treated outside the trial. Preferential treatment of younger or more severely affected patients outside randomisation is possible, but remains unmeasured. It should not be used to assert a benefit in those unseen patients.

Missing primary outcomes were few, but the imputation rules were simple. Carrying forward an earlier disability score assumes no material subsequent change; assigning mRS 4 to surviving patients without either assessment makes a different assumption. The small number limits the likely effect on the overall interpretation, but the data were not literally complete. Quality-of-life outcomes had more missing observations and consequently provide weaker reassurance about patient experience.

Some reporting discrepancies also require care. General anaesthesia is reported as 43.1% in one main-text passage and 41.3% in the supplementary table and discussion. The supplementary as-treated functional-outcome figure is labelled with a smaller population than the main safety analysis. These should not be silently combined. The principal efficacy and safety counts remain clear, and retaining their distinct populations avoids creating a misleading synthetic denominator.^{1,3}

Body of Evidence

The earlier HERMES collaboration established a substantial benefit from thrombectomy in predominantly proximal large-vessel occlusion. That evidence explains the rationale for extending treatment, but its population is not interchangeable with medium- and distal-vessel stroke. Anatomical proximity does not guarantee that the balance of salvageable tissue, spontaneous recanalisation and procedural risk remains constant.⁴

The HERMES M2 analysis was much smaller and contained predominantly proximal M2 occlusions. Its favourable findings generated an important hypothesis, but cannot be treated as definitive evidence for all lesions labelled M2, still less for M3, A2–A3 or P2–P3 occlusions.⁵

DISTAL independently found no improvement in 90-day disability with thrombectomy for its selected medium- or distal-vessel population. This agreement strengthens the case against routine extension of thrombectomy on anatomical grounds alone.⁶

The two 2025 trials still should not be collapsed into one identical experiment. Their vascular eligibility, time windows, disability measures and device requirements differed. ESCAPE-MeVO included proximal M2 occlusions and used a mandated first-line stent retriever; DISTAL addressed a different mix of lesions and allowed operator choice of technique. Their shared negative clinical direction is more informative because it emerged despite those differences, but it does not erase uncertainty in small subgroups.^{1,2}

The accompanying editorial appropriately places randomised evidence above the preceding enthusiasm from observational reports. Its proposed explanations, including selection, treatment delay and technical limitations, are useful research questions. They are not evidence that benefit has already been demonstrated once the trial population is redefined. A plausible explanation for a negative result must be tested rather than allowed to neutralise the result indefinitely.²

For future work, the aim should be a clinically identifiable population and a reproducible treatment strategy with an acceptable safety profile. Trials could examine faster workflows, different devices or carefully specified deficits and perfusion patterns. Those choices need justification before outcomes are known. Repeatedly searching existing data for a favourable subgroup risks producing a treatment rule that works only in the dataset from which it was derived.

For current interpretation, ESCAPE-MeVO establishes that this broadly applied endovascular strategy did not improve recovery over usual care and produced

concerning safety signals. Medical treatment, stroke-unit care and rehabilitation remain active care. The unresolved need is how to improve their outcomes safely, rather than how to preserve a procedural default unsupported by these trials.

Should thrombectomy routinely be offered for medium-vessel occlusion within 12 hours?

ESCAPE-MeVO does not support routine thrombectomy for the population studied: functional outcomes did not improve, and haemorrhage and mortality were concerning. Alternative devices or narrower selection strategies require their own evidence.

References

1. Goyal M, Ospel JM, Ganesh A, Dowlathshahi D, Volders D, Möhlenbruch MA, et al. Endovascular Treatment of Stroke Due to Medium-Vessel Occlusion. *N Engl J Med*. 2025;392(14):1385-1395.
2. Mocco J. Medium- and Distal-Vessel Occlusion - The Limit of Thrombectomy? *N Engl J Med*. 2025;392(14):1440-1442.
3. Goyal M, Ospel JM, Ganesh A, Dowlathshahi D, Volders D, Möhlenbruch MA, et al. Supplementary methods and results to Endovascular Treatment of Stroke Due to Medium-Vessel Occlusion. *N Engl J Med*. 2025.
4. Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet*. 2016;387(10029):1723-1731.
5. Menon BK, Hill MD, Davalos A, Roos YBWEM, Campbell BCV, Dippel DWJ, et al. Efficacy of endovascular thrombectomy in patients with M2 segment middle cerebral artery occlusions: meta-analysis of data from the HERMES Collaboration. *J Neurointerv Surg*. 2019;11(11):1065-1069.
6. Psychogios M, Brehm A, Ribo M, Rizzo F, Strbian D, Rätty S, et al. Endovascular Treatment for Stroke Due to Occlusion of Medium or Distal Vessels. *N Engl J Med*. 2025;392(14):1374-1384.

PEARL

Writing Committee for the PEARL Investigators. Intra-Arterial Alteplase After Successful Endovascular Reperfusion in Acute Stroke: The PEARL Randomized Clinical Trial. JAMA. 2025;334(19):1728–1739.

Introduction

An apparently successful thrombectomy can be followed by substantial disability. The proximal artery may be open while distal branches remain obstructed, or tissue may fail to recover despite restored large-vessel flow. Adjunctive intra-arterial thrombolysis aims to address residual thrombus after the mechanical procedure has reached its intended endpoint. The concept is attractive because the catheter is already in place and the drug can be delivered close to the affected territory.

That convenience should not obscure the additional intervention. A local infusion requires catheter positioning, time and exposure to a fibrinolytic in recently ischaemic tissue. Patients may already have received an intravenous thrombolytic or other antithrombotic drugs. Some have substantial residual non-reperfused territory; others have near-complete or complete angiographic reperfusion. The potential benefit and bleeding risk may differ across these circumstances.

PEARL tested intra-arterial alteplase after successful endovascular reperfusion in anterior circulation large-vessel stroke. It is particularly relevant because preceding intravenous thrombolysis was permitted. The trial found more excellent functional outcomes, but no clear improvement across the full disability distribution and numerically more deaths. Understanding those findings together is essential: the primary endpoint provides evidence of benefit, while the broader outcome pattern determines how confidently that benefit can be translated into routine practice.¹

Synopsis

PEARL was a multicentre, prospective, open-label randomised trial with blinded endpoint assessment at 28 hospitals in China. Patients were recruited from August 2023 to October 2024, with final follow-up in January 2025. Public and institutional grants supported the study, alongside unrestricted grants from two medical-device companies. Alteplase was obtained through hospital formularies. The investigators stated that funders and the drug manufacturer had no role in design, conduct, analysis or publication. An independent committee monitored safety.¹

Eligible adults had an acute ischaemic stroke caused by intracranial internal carotid, M1 or M2 occlusion, NIHSS 6–25 and ASPECTS of at least 6. Randomisation had to occur within 24 hours of last known well after stable reperfusion, eTICI 2b50–3, had been achieved and no further thrombectomy attempts were planned. The reported principal eligibility rule required pre-stroke mRS 0 or 1; the supplementary criteria also allowed greater disability unrelated to neurological disease. Patients already meeting the reperfusion threshold at diagnostic angiography were eligible under the written criteria.^{1,2}

Previous intravenous thrombolysis was allowed, including appropriately selected late-window treatment. Important exclusions included contraindications to alteplase, planned dual antiplatelet treatment within 24 hours, carotid disease requiring stenting, substantial bleeding risks and severe renal dysfunction. The procedure was limited to six retrieval passes, with further mechanical clot manipulation prohibited after randomisation. These restrictions created a selected post-reperfusion population rather than an unrestricted rescue-thrombolysis cohort.^{1,2}

A central web-based system used blocks of four, stratified by centre and preceding intravenous thrombolysis. Independent statisticians generated the allocation sequences. Patients received either intra-arterial alteplase 0.225 mg/kg, maximum 20 mg, over 15 minutes or standard treatment without an intra-arterial fibrinolytic. The infusion catheter was placed proximal to the original target, with more selective positioning left to the operator according to feasibility and risk. Controls were observed for 15 minutes. Both groups then underwent a repeat angiogram.¹

The primary endpoint was excellent functional recovery, mRS 0 or 1 at 90 days. Central trained interviewers, masked to allocation, conducted structured telephone assessments. A separate committee of three certified neurologists adjudicated the final outcome. Secondary endpoints included functional independence, mRS 0–3, the ordinal mRS distribution, early neurological improvement, quality of life, infarct growth and residual hypoperfusion. Principal safety endpoints were symptomatic and any intracranial haemorrhage within 36 hours and mortality by 90 days.¹

The sample-size calculation assumed a control excellent-outcome rate of 40.4% and a 16-percentage-point absolute improvement. A target of 324 patients allowed 80% power and 5% attrition. An interim review after half the planned sample had completed follow-up did not cross the stopping boundary, and recruitment continued. The final primary significance threshold was $P < 0.049$. The primary model accounted for preceding intravenous thrombolysis and pooled study centres. Secondary comparisons were exploratory without multiplicity adjustment.¹

Among 1,879 screened patients, 324 were randomised, 164 to alteplase and 160 to control. One patient in each group lacked the primary follow-up assessment, leaving 163 and 159 for the main efficacy analysis. Multiple-imputation and covariate-adjusted analyses were performed as sensitivity analyses. The per-protocol population contained 160 and 154 patients. Important screening exclusions included refusal of participation and stenting or early dual-antiplatelet requirements.^{1,2}

Median age was 68 years, median NIHSS 15 and median ASPECTS 9. Approximately 31% were women and most were Han Chinese. M1 occlusion accounted for 72% of cases. Previous intravenous thrombolysis had been given to 135 patients, 41.7%, predominantly with alteplase. The report states that the intravenous course was complete by randomisation in those who received it. Approximately 58% had eTICI 2b50 or 2b67 at entry; relatively few had complete reperfusion. Median onset-to-randomisation time was about seven hours.¹

Excellent outcome occurred in 73/163 patients receiving intra-arterial alteplase and 48/159 controls, 44.8% versus 30.2%. The adjusted risk ratio was 1.45 (95% CI 1.08–1.96; $P=0.01$), with an adjusted absolute difference of 14.99 percentage points (95% CI 3.71–26.27). The crude difference from the reported proportions was 14.6 percentage points. The per-protocol and missing-data sensitivity analyses also favoured treatment.^{1,2}

Functional independence occurred in 55.2% versus 48.4%, adjusted risk ratio 1.13 (95% CI 0.92–1.39). The generalised odds ratio for the overall mRS distribution was 1.23 (95% CI 0.92–1.54; $P=0.10$). Neither mRS 0–3, early neurological improvement nor quality of life showed a clear between-group benefit. Infarct growth and residual hypoperfusion also did not differ significantly, although these analyses had appreciable missing data. Angiographic improvement after treatment was assessed post hoc and did not establish a significant between-group difference.^{1,2}

Symptomatic intracranial haemorrhage occurred in 7/164 alteplase patients and 8/160 controls, 4.3% versus 5.0%, adjusted risk ratio 0.85 (95% CI 0.43–1.69). Any intracranial haemorrhage occurred in 32.9% versus 26.9%, adjusted risk ratio 1.22 (95% CI 0.92–1.63). Death by 90 days occurred in 28/164 versus 18/160, 17.1% versus 11.3%, adjusted hazard ratio 1.60 (95% CI 0.88–2.89). Neither safety comparison established a statistically significant difference, but the direction of mortality and overall haemorrhage warrants attention.¹

Critique

Randomising after a stable reperfusion result is an important design strength. The trial isolates the decision to add a fibrinolytic after the operator has decided to stop mechanical treatment. Stratification by preceding intravenous thrombolysis addresses a clinically important variable, and masked central adjudication strengthens the primary outcome assessment. Recruitment continued to the planned sample after interim review, and primary follow-up was nearly complete.

The prespecified primary endpoint was positive and should not be discounted because other outcomes did not reach significance. An excellent recovery is valuable to patients, and the estimate is compatible with a clinically worthwhile benefit. The appropriate interpretation is that PEARL increased the probability of mRS 0–1 under the trial's conditions. That statement is narrower than saying that it improved every aspect of recovery or established a favourable overall benefit–harm balance for routine use.

The lack of a clear ordinal effect is particularly relevant because deaths were more frequent with treatment. A therapy could conceivably increase the number of excellent recoveries while also worsening outcomes in another part of the population. The data do not prove that this occurred causally in PEARL, but the observed distribution makes it an important possibility to examine. A single favourable threshold cannot capture every clinically important feature of that distribution.

Mortality was not the primary endpoint, and the confidence interval includes no difference. It would be excessive to claim that the trial conclusively demonstrated lethal harm. It would be equally inappropriate to dismiss the imbalance as established chance or call the treatment definitively safe. The upper confidence limit permits a substantial increase in the hazard of death. With no proven advantage in the overall disability distribution, that uncertainty has practical weight.

The same caution applies to symptomatic haemorrhage. Fifteen events provide limited precision, and the adjusted interval remains compatible with increased risk despite a point estimate below one. Any intracranial haemorrhage was numerically more frequent with alteplase. The two haemorrhage measures answer different questions, and the absence of a significant difference in one should not be used to erase the direction of the other. Larger studies are needed to characterise uncommon but serious complications adequately.

The trial was open label and did not use a placebo infusion. Observing controls for 15 minutes and repeating angiography is a useful attempt to align the procedural pathway. Nevertheless, the intervention still involved additional microcatheter positioning and

infusion. The trial therefore estimates the effect of that whole package, not a purely pharmacological effect separated from catheter manipulation or flushing. A placebo-controlled design would better isolate the drug contribution, although it would also need to account for the risks of a sham intra-arterial procedure.

Concomitant treatment was not entirely standardised. Tirofiban was used in 26.2% of the alteplase group and 34.3% of controls. The direction in which this might affect the comparison cannot be confidently inferred because use was not randomised and may reflect vascular disease or procedural circumstances. It is not enough to assume that tirofiban was beneficial and therefore made the alteplase estimate conservative. Such an explanation needs evidence in the relevant setting and an understanding of why the additional drug was given.¹

The inclusion of patients receiving intravenous thrombolysis improves relevance to many thrombectomy pathways, but does not establish equal benefit before and after such treatment. The subgroup point estimates differed, with a smaller apparent effect among patients who had received an intravenous fibrinolytic. The interaction test did not establish heterogeneity, and the subgroup sample sizes were limited. Neither selecting only patients without prior thrombolysis nor withholding indicated intravenous treatment is justified by these exploratory results.

The range of reperfusion grades is also central. Most patients still had a meaningful amount of unperfused territory, whereas only a small minority had eTICI 3. Residual distal arterial thrombus provides a plausible target for alteplase, but the trial cannot establish that complete reperfusion identifies either benefit or futility. The post hoc grade-by-treatment analysis was imprecise. Absence of a statistically significant interaction is not proof that the treatment effect is identical across grades.

The mechanistic evidence was less persuasive than the primary clinical result. Neither infarct growth nor 24-hour hypoperfusion showed a clear treatment difference. These findings do not disprove benefit, because measurements were incomplete and may not capture microvascular flow adequately. They do mean that improved microcirculation remains a proposed explanation rather than a demonstrated mediator. Any visible angiographic improvement could also reflect distal macroscopic recanalisation rather than capillary no-reflow reversal.

Missing imaging and quality-of-life data further constrain interpretation. The main table identifies 60 missing infarct-growth assessments and 70 missing perfusion measurements, with poor image quality, serious illness or death among the reasons. The narrative gives a different count for completed perfusion follow-up, which needs

reconciliation. Quality-of-life assessment also excluded deaths and had additional missing observations. These are not necessarily random omissions, so apparently similar results among measured survivors do not establish comparable outcomes for the full randomised population.¹

The sample size assumed a large benefit and the actual control excellent-outcome rate was below the anticipated value. Sensitivity models adding age and sex remained favourable, but two had P values just below the final 0.049 threshold and lower confidence limits close to one. This does not overturn the primary analysis. It does show that the apparent strength and precision of the finding depend somewhat on the modelling approach, making independent replication more valuable.²

A post hoc analysis combined severe disability and death into one mRS category and obtained a conventionally significant ordinal result only in the per-protocol population. That analysis changes both the outcome structure and the population. It cannot replace the prespecified randomised analysis or resolve concern about mortality by making death statistically indistinguishable from severe disability. It is appropriately hypothesis-generating, not confirmatory evidence of a broad functional benefit.²

Generalisability is constrained by the exclusion of large infarcts, early dual-antiplatelet requirements and stenting, and by limited demographic diversity. These exclusions remove common difficult cases in which clinicians may be particularly tempted to add rescue treatment. The trial also used a specific dose cap of 20 mg, which should not be confused with other intra-arterial alteplase regimens. An adjunct that appears feasible in selected successful procedures is not automatically appropriate after a complicated or unsuccessful thrombectomy.

Body of Evidence

CHOICE provided the earlier placebo-controlled signal for intra-arterial alteplase after successful thrombectomy. It reported more excellent outcomes, but ended before its planned recruitment was complete because of logistical difficulties. PEARL is an important further test of the same general strategy in another health-care setting, with a larger sample. The regimens were not identical: PEARL capped alteplase at 20 mg, whereas CHOICE used a 22.5 mg maximum.³

ANGEL-TNK also found more patients with mRS 0–1 after adjunctive intra-arterial thrombolysis, using tenecteplase rather than alteplase. It excluded preceding intravenous thrombolysis and enrolled a selected late-window population. Like PEARL, its positive primary dichotomy was not accompanied by a clear overall disability shift. The

agreement therefore concerns a particular outcome signal, not proof that the two drugs are interchangeable or that broader functional and safety uncertainty has been resolved.⁴

POST-TNK tested a lower tenecteplase dose after near-complete or complete reperfusion and did not significantly improve its primary excellent-outcome endpoint. POST-UK similarly did not establish a primary benefit from intra-arterial urokinase in a population with near-complete or complete reperfusion. These studies differ from PEARL in agent, dose, angiographic eligibility and preceding thrombolysis. Their results should contribute to the overall uncertainty rather than be dismissed as the wrong dose after their outcomes are known.^{5,6}

ATTENTION-IA addressed posterior circulation occlusion and also failed to establish a significant primary benefit from adjunctive intra-arterial tenecteplase. It is not a direct comparator for PEARL's anterior circulation population. Its relevance is to the limits of a broad class claim: successful recanalisation and an available microcatheter do not, by themselves, define a population with proven benefit from another fibrinolytic.⁷

The accompanying PEARL editorial recognises the promise of the primary result while calling for larger confirmatory trials. Its suggestions about residual distal thrombus, dose selection and procedural refinements are reasonable research directions. They remain hypotheses. In particular, the proposal that patients with incomplete reperfusion benefit more than those with complete reperfusion has not been established by a definitive interaction in PEARL.⁸

The evidence available by the end of 2025 therefore supports continued investigation of adjunctive intra-arterial thrombolysis, with several positive excellent-outcome signals but inconsistent broader functional results. Pooling different agents and selection rules may improve numerical precision while obscuring the actual treatment decision. Further work should preserve those differences, assess the complete disability distribution and provide reliable safety estimates, especially after preceding intravenous thrombolysis.

PEARL advances the field by showing a positive primary result in a population that included patients already exposed to intravenous fibrinolysis. It also makes clear what remains unresolved: the clinical importance of the mortality imbalance, the optimal recipient, the best drug and dose, and whether the primary signal translates into a reproducible improvement in overall patient outcome. Those uncertainties should accompany, rather than be hidden behind, the positive primary P value.

Should intra-arterial alteplase routinely be added after successful anterior circulation thrombectomy?

PEARL increased excellent recovery, but broader disability benefit was uncertain and mortality was numerically higher. It supports further confirmatory study rather than routine treatment after every successful procedure.

References

1. Writing Committee for the PEARL Investigators. Intra-Arterial Alteplase After Successful Endovascular Reperfusion in Acute Stroke: The PEARL Randomized Clinical Trial. *JAMA*. 2025;334(19):1728–1739.
2. Writing Committee for the PEARL Investigators. Supplementary methods and results to Intra-Arterial Alteplase After Successful Endovascular Reperfusion in Acute Stroke: The PEARL Randomized Clinical Trial. *JAMA*. 2025.
3. Renú A, Millán M, San Román L, Blasco J, Martí-Fàbregas J, Terceño M, et al. Effect of Intra-arterial Alteplase vs Placebo Following Successful Thrombectomy on Functional Outcomes in Patients With Large Vessel Occlusion Acute Ischemic Stroke: The CHOICE Randomized Clinical Trial. *JAMA*. 2022;327(9):826-835.
4. Miao Z, Luo G, Song L, Sun D, Chen W, Yao X, et al. Intra-arterial Tenecteplase for Acute Stroke After Successful Endovascular Therapy: The ANGEL-TNK Randomized Clinical Trial. *JAMA*. 2025;334(7):582-591.
5. Huang J, Yang J, Liu C, Li L, Yang D, Guo C, et al. Intra-Arterial Tenecteplase Following Endovascular Reperfusion for Large Vessel Occlusion Acute Ischemic Stroke: The POST-TNK Randomized Clinical Trial. *JAMA*. 2025;333(7):579-588.
6. Liu C, Guo C, Li F, Yu N, Huang J, Peng Z, et al. Intra-Arterial Urokinase After Endovascular Reperfusion for Acute Ischemic Stroke: The POST-UK Randomized Clinical Trial. *JAMA*. 2025;333(7):589-598.
7. Hu W, Tao C, Wang L, Chen Z, Li D, Chen W, et al. Intra-arterial tenecteplase after successful endovascular recanalisation in patients with acute posterior circulation arterial occlusion (ATTENTION-IA): multicentre randomised controlled trial. *BMJ*. 2025;388:e080489.
8. Seitz A, Leslie-Mazwi TM. Intra-Arterial Thrombolysis After Thrombectomy—Does the Story Have an Epilogue? *JAMA*. 2025;334(19):1713–1715.

A2B

**Walsh TS, Parker RA, Aitken LM, McKenzie CA, Emerson L, Boyd J, et al.
Dexmedetomidine- or Clonidine-Based Sedation Compared With Propofol in
Critically Ill Patients: The A2B Randomized Clinical Trial. JAMA. 2025;334(1):32–45.**

Introduction

Sedation can make invasive ventilation tolerable while also delaying recovery through excessive suppression of consciousness, respiratory drive and mobilisation. The practical objective is therefore adequate comfort with the least sedation required by the clinical situation. Selecting a drug is only one part of that process. Analgesia, regular assessment, explicit sedation targets and timely recognition of readiness for extubation can all influence the apparent performance of a sedative.

Propofol provides rapidly adjustable sedation over a broad range of depths. Dexmedetomidine and clonidine act through α_2 -adrenergic receptors and can produce a more arousable state with less respiratory depression. These properties create a plausible route to earlier liberation and less acute brain dysfunction. However, sympatholysis and bradycardia can limit their use, and additional agents are frequently necessary when deeper sedation is required. A physiological advantage does not automatically produce a better clinical outcome.¹

Much of the early enthusiasm for dexmedetomidine arose from comparisons with benzodiazepines. The clinically relevant question changes when the alternative is propofol used within a light-sedation strategy. Clonidine adds another uncertainty: its lower acquisition cost and familiar pharmacology make it attractive, but neither membership of the same drug class nor a neutral comparison with usual care establishes equivalence to dexmedetomidine.

A2B addressed these questions in a large pragmatic trial of early sedation strategies. It compared dexmedetomidine-based and intravenous clonidine-based sedation with propofol-based care, and planned a direct non-inferiority comparison only if the initial superiority testing justified proceeding. The trial therefore needs to be read both as a clinical comparison and as a sequence of statistical questions whose later stages were conditional.

Synopsis

A2B was an open-label, three-group randomised trial conducted in 41 UK ICUs between December 2018 and October 2023. The National Institute for Health and Care Research funded the study and had no role in its conduct, analysis or publication decisions. Recruitment was disrupted by the COVID-19 pandemic. The principal report was published online on 19 May 2025 and subsequently corrected; the corrected report is used here.^{1,2}

Eligible adults were receiving propofol within 48 hours of starting invasive ventilation in the ICU and were expected to require at least another 24 hours of ventilation, with a total duration of at least 48 hours. Important exclusions included acute brain injury, concern about hypoxic brain injury after cardiac arrest, status epilepticus, continuous therapeutic neuromuscular paralysis, relevant neuromuscular disease, fulminant hepatic failure, untreated advanced heart block and sustained marked bradycardia. Patients expected to die within 24 hours or receiving end-of-life care were excluded.

Central randomisation used concealed allocation, variable-sized blocks and stratification by centre. Patients were allocated equally to dexmedetomidine, clonidine or propofol strategies. Treatment was unblinded. The intention was to introduce the allocated α_2 agonist promptly and reduce propofol, while allowing supplementary propofol if the maximum agonist dose was insufficient or adverse effects limited titration.

The suggested starting infusion was 0.7 $\mu\text{g/kg/hour}$ for dexmedetomidine and 1 $\mu\text{g/kg/hour}$ for clonidine, with lower starting doses where cardiovascular instability warranted caution. Maximum doses were 1.4 and 2 $\mu\text{g/kg/hour}$, respectively. Both drugs were administered intravenously. The usual target was a Richmond Agitation-Sedation Scale score between -2 and +1, although clinicians could request deeper sedation when indicated. Opioid analgesia was determined clinically and benzodiazepines were discouraged. Treatment continued until successful extubation, death, transfer outside participating care or day 28.

The primary outcome was time to successful extubation, requiring a subsequent 48-hour period of spontaneous breathing without mechanical ventilation. Patients receiving non-invasive support needed to meet the specified low-support definition, including no more than 5 cm H₂O continuous positive airway pressure. Death was treated as a competing event in the primary subdistribution hazard model. Secondary outcomes included mortality to 180 days, ICU discharge, sedation quality, delirium, cardiovascular events and longer-term patient-reported outcomes.

The testing hierarchy first compared each α_2 -agonist strategy with propofol for superiority, using a 2.5% significance level for each comparison. The planned clonidine-versus-dexmedetomidine non-inferiority assessment was conditional on the preceding tests. The recruitment target was reduced during the pandemic from 1,737 to 1,437, retaining high stated power for the anticipated superiority effect but reducing power for the subsequent non-inferiority question.

There were 1,438 randomisations, including one duplicate, and 1,404 patients entered the primary analysis: 457 dexmedetomidine, 476 clonidine and 471 propofol. Exclusions included consent problems, erroneous or duplicate randomisation and fourteen patients at a site with a serious protocol breach. Median time from starting ventilation to randomisation was approximately 21 hours. About two-thirds had sepsis, and substantial cardiovascular dysfunction was common. Most patients were deeply sedated immediately before allocation.

Neither intervention significantly improved the primary outcome. The subdistribution hazard ratio for successful extubation was 1.09 with dexmedetomidine versus propofol (95% CI 0.96–1.25; $P=0.20$) and 1.05 with clonidine versus propofol (95% CI 0.95–1.17; $P=0.34$). Descriptive median times were 136, 146 and 162 hours, respectively. Prespecified sensitivity analyses supported the primary findings. Because neither superiority test was positive, the later hypotheses in the hierarchy were not formally tested.

Mortality at 180 days was 132/457 (29%) with dexmedetomidine, 145/476 (30%) with clonidine and 141/471 (30%) with propofol. Corresponding mortality hazard ratios versus propofol were 0.98 (95% CI 0.77–1.24) and 1.04 (95% CI 0.82–1.31). There was no clear difference in ICU discharge or measured delirium outcomes. The median time to first achieve a RASS score of at least -2 was two twelve-hour nursing shifts in each group.

The interventions changed drug exposure without producing exclusive single-agent sedation. Dexmedetomidine and clonidine were each given for a median of four days. Propofol was also used on 77% and 76% of days in those groups, respectively, at lower doses than in the propofol group. Opioid doses were similar. These findings describe combinations led by different primary agents.

Agitation was more frequent with both α_2 -agonist strategies: rate ratios were 1.54 (95% CI 1.21–1.97) for dexmedetomidine and 1.55 (95% CI 1.22–1.97) for clonidine versus propofol. Severe bradycardia occurred in approximately 33% of patients in each agonist group and 20% of the propofol group. Event-rate ratios were 1.62 (95% CI 1.36–1.93) and

1.58 (95% CI 1.33–1.88), respectively. No clinically important advantage was apparent in the longer-term outcomes among respondents.^{1,2}

Critique

A2B does not support routinely replacing propofol-based sedation with either α_2 -agonist strategy to accelerate extubation. Its pragmatic design, broad UK participation and clinically relevant comparator make that conclusion useful. The result also directly informs clonidine practice, for which large randomised comparisons had been limited. However, the trial does not establish that all three drugs are interchangeable, nor that dexmedetomidine or clonidine has no role in selected patients.

The descriptive medians should not become the headline treatment effect. Subtracting 136 from 162 hours suggests a 26-hour advantage for dexmedetomidine, but that arithmetic is not the prespecified competing-risk comparison. The median estimates and the primary model handle event timing and competing death differently. The model did not establish superiority, and its confidence interval includes both no benefit and a potentially useful increase in the rate of successful extubation. A claim of a proven one-day reduction would overstate the result.

The hierarchical design is equally important. Failure to demonstrate that either agonist was superior to propofol prevented formal progression to the planned non-inferiority comparison. Similar observed outcomes between clonidine and dexmedetomidine cannot substitute for that test. The trial may make intravenous clonidine a more evidence-informed option, but it does not prove that clonidine preserves an established dexmedetomidine benefit within an acceptable margin. Cost arguments should not be built on an equivalence claim the trial never established.

The intervention is best understood as a strategy. Substantial supplementary propofol exposure limits inference about the isolated pharmacological effects of either agonist. It does not necessarily represent a failed trial: patients often need combinations because agitation, shock and changing sedation requirements prevent single-agent treatment. A strategy that cannot be delivered without adjuncts in usual practice should be evaluated in that form. Conversely, the results cannot answer whether a carefully selected population receiving agonist monotherapy would respond differently.

Actual sedation depth helps explain the clinical result. All groups achieved light-sedation targets relatively promptly, and the time to first adequate arousal was similar. The hypothesised pathway from agonist allocation to greater wakefulness and then earlier extubation was therefore not strongly separated. Deep sedation was still requested on a

substantial minority of days in every group. This may reflect appropriate clinical indications, established practice or both; the data cannot identify the contribution of each.

Open treatment allocation matters because successful extubation is partly clinician-mediated. Clinicians may be comfortable extubating during an α_2 -agonist infusion but prefer to stop propofol first. Decisions about spontaneous breathing trials, extubation and post-extubation support could therefore differ despite common general guidance. The accompanying editorial also highlights uncertainty about standardisation of non-invasive ventilation after extubation. A strict endpoint requiring minimal support may classify appropriate early extubation to non-invasive ventilation differently from complete liberation.³

These concerns do not establish the direction or magnitude of bias. An open-label design may favour a familiar intervention, a hoped-for new treatment or neither, depending on local behaviour. The central allocation process protects the initial comparison, and sensitivity analyses provide reassurance. Nevertheless, objectively measured outcomes such as mortality deserve separate attention from treatment decisions made by unblinded staff.

Bradycardia is a clinically relevant finding, rather than merely a pharmacological curiosity. It was more frequent despite exclusion of patients with sustained severe bradycardia before randomisation. The event-rate ratios derive from repeated observations and should not be confused with a simple relative risk of ever having an event. The approximately 33% versus 20% patient-level prevalence provides a complementary perspective. Whether an episode required intervention, restricted dosing or affected recovery is a further question; the trial does not turn every recorded episode into a serious clinical injury.

Agitation also complicates a simple account of more comfortable or cooperative sedation. Both agonist groups had higher measured agitation rates without a clear reduction in pain behaviours, unnecessary deep sedation or opioid exposure. Clinician familiarity and dose limitation may have contributed, but those explanations remain interpretations. They should not be used to dismiss an unfavourable outcome that arose under the tested strategy. Choice of sedative still requires attention to both wakefulness and distress.

The mortality comparison is reassuring against a very large average difference, but its intervals remain compatible with clinically important benefit or harm. A2B was designed around extubation, not mortality equivalence. Prespecified categorical subgroup

analyses did not identify a convincing treatment interaction. A weak continuous-age interaction for extubation with dexmedetomidine is insufficient to define an age threshold for prescribing, particularly alongside multiple analyses and no corresponding mortality interaction.

Applicability is strongest for general ICU patients already receiving propofol early in ventilation. Acute brain injury, post-arrest concern about hypoxic injury, continuous paralysis and several neuromuscular conditions were excluded. These exclusions remove important groups in whom sedation depth and extubation are governed by different considerations. The high prevalence of shock is relevant to real UK practice, but may also constrain agonist dosing. Extrapolation to prolonged use later in the ICU course, withdrawal management or sedation during non-invasive ventilation requires other evidence.

Longer-term outcomes are a weaker basis for reassurance. Questionnaire completion was limited: the 180-day anxiety and depression assessment included 69, eighty and 85 patients across the three groups. Similar scores among respondents do not demonstrate absence of a long-term effect in all survivors, particularly if response depends on disability or cognitive recovery. Mortality follow-up and questionnaire follow-up should not be described as equally complete.

Finally, the published correction concerned the selection of participants used to summarise several baseline variables, including age, sex, weight and pre-randomisation sedation assessments. The corrected version is the appropriate source for those characteristics. This does not invalidate the outcome comparison, but it reinforces the need to distinguish a correction to baseline description from evidence that the clinical endpoint changed. Any remaining discrepancies between corrected tables and narrative summaries should be resolved editorially before final publication.

Body of Evidence

MENDS provided an early signal that sedative choice could influence acute brain dysfunction. In 106 ventilated patients, dexmedetomidine produced more days alive without delirium or coma than lorazepam, with medians of seven versus three days. It also improved time near the sedation target. This was a small comparison against a benzodiazepine, however, and did not establish a survival advantage. Its findings support avoiding an assumption that all sedative comparators carry the same baseline risk.⁴

SEDCOM similarly compared dexmedetomidine with midazolam. Time within the target sedation range was similar, while delirium prevalence was lower and median time to

extubation shorter with dexmedetomidine. Bradycardia was substantially more frequent. The trial therefore anticipated both an efficacy signal against a benzodiazepine and the cardiovascular trade-off seen in later studies. It did not answer whether the same benefit would occur against well-managed propofol.⁵

The paired MIDEX and PRODEX trials help separate those questions. Dexmedetomidine was non-inferior for time at target light-to-moderate sedation against both midazolam and propofol. Ventilation duration was shorter against midazolam, but the propofol comparison did not show a significant reduction. Improved ability to communicate pain did not translate into clear differences in ICU stay or mortality. This comparator dependence is consistent with A2B's failure to demonstrate superior liberation against propofol.⁶

SPICE III tested early dexmedetomidine as the primary sedative in a much larger population. Among 3,904 patients in the modified intention-to-treat analysis, 90-day mortality was 29.1% in both groups, with an adjusted risk difference of 0.0 percentage points (95% CI -2.9 to 2.8). Additional sedatives were frequently needed, and bradycardia and hypotension were more common with dexmedetomidine. A2B therefore did not introduce the problem of adjunctive treatment; it further examined a practical limitation already apparent in a major preceding trial.⁷

MENDS2 is particularly relevant because it compared dexmedetomidine directly with propofol in ventilated adults with sepsis using a double-blind design. In 422 analysed patients, adjusted days alive without delirium or coma were 10.7 versus 10.8, and neither ventilator-free days nor 90-day mortality differed clearly. Six-month cognitive outcomes were also similar within the available data. Together with the high proportion of sepsis in A2B, these findings weaken the argument for routinely preferring dexmedetomidine on the expectation of a major delirium or survival benefit in sepsis.⁸

The evidence therefore supports a distinction between reducing exposure to benzodiazepine-heavy sedation and identifying a universally superior replacement for propofol. A2B adds a large direct assessment of intravenous clonidine and confirms that changing the primary drug does not guarantee earlier liberation when sedation targets and background care are similar. It leaves room for individual choice based on required depth, arousability, cardiovascular tolerance and the patient's changing clinical state. That choice should be framed around achievable treatment goals, not an unproven class-wide advantage.

Should dexmedetomidine or clonidine routinely replace propofol to shorten ventilation?

A2B does not support routine replacement for this purpose: neither strategy significantly accelerated successful extubation, and both increased bradycardia and agitation. It also did not establish clonidine's non-inferiority to dexmedetomidine.

References

1. Walsh TS, Parker RA, Aitken LM, McKenzie CA, Emerson L, Boyd J, et al. Dexmedetomidine- or Clonidine-Based Sedation Compared With Propofol in Critically Ill Patients: The A2B Randomized Clinical Trial. *JAMA*. 2025;334(1):32-45.
2. Walsh TS, Parker RA, Aitken LM, McKenzie CA, Emerson L, Boyd J, et al. A2B supplemental online content. *JAMA*. 2025. doi:10.1001/jama.2025.7200.
3. Qian ET, Rice TW. The Search for the Single Best Sedative. *JAMA*. 2025;334(1):26–28.
4. Pandharipande PP, Pun BT, Herr DL, Maze M, Girard TD, Miller RR, et al. Effect of sedation with dexmedetomidine vs lorazepam on acute brain dysfunction in mechanically ventilated patients: the MENDS randomized controlled trial. *JAMA*. 2007;298(22):2644-2653.
5. Riker RR, Shehabi Y, Bokesch PM, Ceraso D, Wisemandle W, Koura F, et al. Dexmedetomidine vs midazolam for sedation of critically ill patients: a randomized trial. *JAMA*. 2009;301(5):489-499.
6. Jakob SM, Ruokonen E, Grounds RM, Sarapohja T, Garratt C, Pocock SJ, et al. Dexmedetomidine vs midazolam or propofol for sedation during prolonged mechanical ventilation: two randomized controlled trials. *JAMA*. 2012;307(11):1151-1160.
7. Shehabi Y, Howe BD, Bellomo R, Arabi YM, Bailey M, Bass FE, et al. Early Sedation with Dexmedetomidine in Critically Ill Patients. *N Engl J Med*. 2019;380(26):2506-2517.
8. Hughes CG, Mailloux PT, Devlin JW, Swan JT, Sanders RD, Anzueto A, et al. Dexmedetomidine or Propofol for Sedation in Mechanically Ventilated Adults with Sepsis. *N Engl J Med*. 2021;384(15):1424-1436.

B-Free

Spence J, Devereaux PJ, Lee SF, D'Aragon F, Avidan MS, Whitlock RP, et al; B-Free Investigators and the Canadian Perioperative Anesthesia Clinical Trials Group. Benzodiazepine-Free Cardiac Anesthesia for Reduction of Postoperative Delirium: A Cluster Randomized Crossover Trial. JAMA Surg. 2025;160(3):286–294.

Introduction

Delirium after cardiac surgery arises from an interaction between pre-existing vulnerability and perioperative insults. Anaesthetic drugs are potentially modifiable contributors, but their effects cannot easily be separated from haemodynamic instability, surgical complexity, inflammation, pain and postoperative sedation. An inexpensive change in routine anaesthesia would be attractive if it prevented delirium without impairing the quality or safety of anaesthesia. The size of any benefit matters: cardiac surgery is common enough for a small individual effect to have an important population consequence.¹

Benzodiazepines occupy an uncertain position in this discussion. Trials of prolonged intensive care sedation have favoured dexmedetomidine over lorazepam or midazolam for several measures of acute brain dysfunction.^{2,3} These findings do not directly establish the effect of a limited intraoperative dose. Exposure over several days, depth of sedation, drug accumulation and the alternative sedative all differ from contemporary cardiac anaesthesia. Conversely, habitual perioperative use cannot demonstrate that benzodiazepines are necessary or harmless.

One reason for administering a benzodiazepine during cardiac surgery is concern about awareness under anaesthesia. Its amnestic properties may appear reassuring when haemodynamic instability limits other agents. However, whether a benzodiazepine is required to prevent awareness is distinct from whether avoiding it reduces delirium. Both questions require appropriate outcome ascertainment. A patient who does not spontaneously report recall has not necessarily undergone an adequate assessment for awareness.

B-Free approached the problem as a hospital-policy question. Rather than recruiting selected individuals to a tightly controlled drug regimen, it alternated restrictive and liberal intraoperative policies across entire cardiac surgical centres. The trial therefore tested what happens when a service changes its default practice while retaining clinical discretion. Its large sample and embedded design make it particularly relevant to

institutional decisions, provided its result is not mistaken for a definitive pharmacological comparison of any benzodiazepine exposure versus none.¹

Synopsis

B-Free was a pragmatic, patient- and assessor-masked, multiperiod cluster randomised crossover trial at 20 North American cardiac surgical centres. Eighteen centres were in Canada and two in the United States. Recruitment occurred from November 2019 to December 2022, with two pauses related to the COVID-19 pandemic. Centres had to secure agreement from at least 95% of their cardiac anaesthetists to follow the allocated policy and routinely assess patients for delirium at least every 12 hours.^{1,4}

All adults undergoing open cardiac surgery at a participating hospital during an active period were eligible for inclusion. Percutaneous procedures and repeat operations during the same admission were excluded. Individual consent was waived; patients received information about the evaluation and could opt out of having their anonymised data included. This design captured a broad clinical population and minimised selective recruitment after allocation. One eligible patient requested removal of their data.

An independent statistician randomised hospitals to restrictive or liberal policies in four-week periods. Randomisation was blocked in pairs of periods to reduce the influence of time trends. Hospitals completed 12–18 periods, with the next allocation communicated during the final week of the preceding period. Anaesthetists knew the policy through email and signs on anaesthetic machines. Patients and statisticians were masked; postoperative clinical staff responsible for delirium assessment were not informed of the trial allocation or that the trial was occurring.

The restrictive policy specified no routine intraoperative benzodiazepines. Exceptions were permitted when the anaesthetist considered them necessary, including particular clinical circumstances such as withdrawal, seizures or haemodynamic instability. The liberal policy specified at least 0.03 mg/kg of midazolam equivalent, using standard ideal weights of 60 kg for women and 70 kg for men. Benzodiazepine choice was not restricted, and exceptions were allowed for contraindications. Other anaesthetic management, preoperative benzodiazepine use and postoperative sedation were not assigned by the trial.

The primary endpoint was delirium detected during routine care within 72 hours of surgery, using the Confusion Assessment Method–Intensive Care Unit or the Intensive Care Delirium Screening Checklist. Secondary outcomes were intensive care and hospital length of stay and in-hospital mortality. Intraoperative awareness was collected through

spontaneous patient reports. The study did not use a formal interview programme to ascertain this rare outcome.

The original design sought more than 80% power to detect a 15% relative reduction in delirium, with assumptions about hospital volume and correlation among patients within a centre. Pandemic-related reductions in case volume and a higher observed intracluster correlation prompted sample-size reassessment using aggregate data from 11,222 patients without unmasking the treatment effect. Additional periods were added where feasible: nine hospitals completed 18, two completed 17 and nine completed 12 periods. The ordinary supplement describes simulations used to evaluate this adaptation.⁴

The primary analysis followed allocation to hospital policy. Mixed models accounted for clustering, period and the changing strength of within-centre correlation over time. Adjustment included age, sex, emergency surgery, heavy alcohol consumption and home benzodiazepine use. Patients with no recorded delirium assessment during the first 72 hours were assumed not to have delirium in the primary analysis. Additional analyses excluded these patients. Several other sensitivity and subgroup analyses were post hoc, and there was no adjustment for multiple testing.

The analysis included 19,768 patients: 9,827 during restrictive periods and 9,941 during liberal periods. Mean age was 65 years and 73.5% were men. Emergency surgery accounted for 7.8%, and 95.1% underwent cardiopulmonary bypass. Home benzodiazepine use and heavy alcohol use were relatively uncommon. Approximately 16% received benzodiazepines during the 24 hours before surgery. The groups were similar in measured patient characteristics and operative details.

The policy produced a large separation in intraoperative exposure. Benzodiazepines were administered to 9.1% of patients during restrictive periods and 93.2% during liberal periods. Among recipients, the mean dose was approximately 4.1 mg of midazolam equivalent. Anaesthetists adhered to the allocated policy in 90.9% and 93.2%, respectively. Some non-benzodiazepine induction doses were higher during restrictive periods, but broad anaesthetic practice was otherwise similar. Postoperative benzodiazepine use was recorded in 10.3% versus 11.1%, and dexmedetomidine use was 14.0% in both groups.

Delirium was assessed at least once in 95.6% of patients. It occurred in 1,373 patients during restrictive periods (14.0%) and 1,485 during liberal periods (14.9%). The adjusted odds ratio was 0.92 (95% confidence interval 0.84–1.01; $P=0.07$). Thus, the prespecified analysis did not establish superiority of the restrictive policy. Its confidence interval

remained compatible with a modest reduction in the odds of delirium as well as essentially no difference.¹

Intensive care length of stay was approximately 3.5 days in both groups in the adjusted analysis, and hospital stay was also similar. In-hospital mortality was 3.0% versus 2.7%, with an adjusted odds ratio of 1.13 (95% confidence interval 0.95–1.34). No patient spontaneously reported intraoperative awareness. Prespecified analyses did not establish different effects according to age, heavy alcohol use or home benzodiazepine use.

Post hoc analyses were more favourable to restriction. Among patients treated according to their allocated intraoperative policy, delirium occurred in 13.7% versus 14.8%, with an adjusted odds ratio of 0.90 (95% confidence interval 0.82–0.99). Excluding patients in the restrictive group who had received preoperative benzodiazepines also produced estimates favouring restriction. The count of assessments documenting delirium was lower during restrictive periods. In contrast, excluding patients without any delirium assessment left the main finding essentially unchanged, with an adjusted odds ratio of 0.92 (0.85–1.01). These results were exploratory rather than replacement primary analyses.

Critique

B-Free is a substantial achievement in pragmatic trial delivery. Nearly 20,000 patients were included across 20 centres without relying on individual recruitment during a stressful perioperative period. The intervention altered ordinary institutional practice, and its effect on drug exposure was large. Repeated crossover allowed each centre to contribute experience under both policies, reducing confounding by stable differences in local surgery, anaesthesia and postoperative care. The study answers an implementation question that a small single-centre drug trial could not address as directly.

The unit of randomisation nevertheless remains important. Thousands of patients do not provide the same independent information as thousands of individually randomised allocations. Outcomes within a centre and period are correlated, and care changes over time. The investigators modelled these features and responded to a higher-than-expected correlation by extending the trial. The blinded adaptation was described transparently and did not simply continue recruitment until a favourable treatment effect appeared. Its success still depends on the adequacy of the assumed correlation structure and on how the pandemic affected case mix and practice across periods.

An institutional crossover also cannot erase learning or behavioural carryover. Anaesthetists experienced both strategies repeatedly, and the ease of avoiding benzodiazepines might influence later practice. However, the sustained exposure contrast argues against major loss of policy separation. The short intraoperative exposure has little direct pharmacological carryover to different patients in a later period; organisational behaviour is the more relevant concern. Period adjustment and paired allocation help address time trends, without guaranteeing that every evolving feature of care is captured.

The primary result is often vulnerable to overinterpretation in both directions. A P value of 0.07 does not establish that benzodiazepine restriction works, but neither does it establish that benzodiazepines have no contribution to delirium. The point estimate favoured restriction and the interval included a potentially useful modest effect. The observed crude difference was approximately one percentage point. The trial was designed to detect a larger relative reduction, so the distinction between failure to demonstrate its intended benefit and proof of no benefit is clinically relevant.

Equally, a policy of routine restriction cannot be described as superior solely because its point estimate is favourable. Mortality and length of stay did not improve clearly. There is no confirmatory evidence here of better long-term cognition, quality of life or functional independence. The evidence supports an informed choice about a default intraoperative practice, while leaving uncertainty about the magnitude of delirium prevention and the best strategy for particular patients.

Delirium ascertainment is a central limitation. Routine use of validated instruments is a strength, and postoperative assessors were not informed of allocation. However, the trial measured detected delirium within 72 hours, not every episode during hospitalisation. Fluctuating or hypoactive delirium may be missed, and sedation or inability to assess can affect observation. The supplement shows that assessment became less frequent in later daily windows. A short observation period also misses later-onset delirium and cannot adequately characterise duration or persistent cognitive effects.⁴

The assumption that patients without any assessment had no delirium is particularly strong. The proportions without assessment were small and similar between groups, and a complete-case analysis gave a similar estimate. Nevertheless, missingness need not be benign simply because its frequency is balanced. Patients dying early or remaining deeply sedated may be less assessable. The substantially lower death counts in the assessed-only sensitivity population indicate that the excluded population was clinically

different. This supports caution about interpreting the recorded delirium rate as the true incidence of acute brain dysfunction.

The trial also does not compare complete benzodiazepine avoidance with exposure. Approximately one-sixth of patients received a preoperative benzodiazepine, postoperative use was allowed and some restrictive-period patients received an intraoperative dose. These features may attenuate a biological contrast, but they are not merely technical flaws: they are part of the policy being tested. A broad “benzodiazepine-free” pathway covering the entire perioperative period could have a different effect, including consequences for withdrawal, anxiety and the use of replacement drugs. Its efficacy and safety cannot be inferred by relabelling the current comparison.

The post hoc analyses excluding non-adherent patients do not solve this problem. Exceptions were chosen by clinicians for clinical reasons, and those reasons can also predict delirium. Removing such patients breaks the protection afforded by randomised policy allocation. Excluding preoperatively exposed patients from the restrictive group while retaining them in the liberal group creates an additional asymmetry. Adjustment for recorded baseline factors cannot ensure that the resulting groups are exchangeable. These analyses support further study of exposure and adherence, but are not independent proof of efficacy among patients who follow the policy.

Likewise, fewer positive delirium assessments may suggest less delirium burden, but the count is influenced by the number and timing of assessments, length of stay, sedation and survival. It is not automatically equivalent to fewer days of delirium measured systematically. This endpoint was post hoc, and several exploratory comparisons were performed without multiplicity correction. A marginally favourable confidence interval should therefore carry less weight than the prespecified primary analysis.

Awareness deserves a separate interpretation. No spontaneous reports across a large sample offers some reassurance about obvious clinically reported events, but the sensitivity of passive reporting is limited. A zero count cannot establish equivalence between policies when systematic interviews were not performed. Some patients may not volunteer an experience, may report it later, or may require structured questioning to distinguish recall from dreams and other perioperative experiences. B-Free consequently does not prove that omitting benzodiazepines is incapable of increasing awareness, nor that giving them is necessary to prevent it.

The accompanying commentary suggested that benzodiazepines were unlikely to be a major culprit in contemporary cardiac-surgery delirium.⁵ The modest doses and

multifactorial nature of delirium make that interpretation plausible, but ruling benzodiazepines out entirely goes beyond the trial. The commentary's broader discussion of preoperative anxiolysis also extends beyond the assigned intraoperative policy. Treatment timing must remain explicit when applying either the trial or its editorial interpretation.

Generalisation is strongest to services resembling the participating centres: predominantly Canadian, established cardiac surgical programmes with routine delirium screening and clinicians willing to alternate policies. The result is less direct for high-dose benzodiazepine anaesthesia, prolonged intensive care sedation, other operations or settings with different monitoring and postoperative care. The trial received peer-reviewed public and research funding, with no funder role in conduct or interpretation reported. Its main value is a credible estimate of the effect of a feasible service-level change, with residual uncertainty that should remain visible in practice decisions.

Body of Evidence

The MENDS trial provides an important example of why sedative context matters. It randomised 106 mechanically ventilated medical and surgical intensive care patients to dexmedetomidine or lorazepam for up to 120 hours. Dexmedetomidine produced more days alive without delirium or coma, with medians of seven versus three days, and less coma. The intervention was sustained sedation with alternative agents, not withdrawal of a small intraoperative benzodiazepine dose. Its positive composite brain-dysfunction result should therefore inform intensive care sedation without being treated as a direct prediction of B-Free.²

SEDCOM similarly compared prolonged dexmedetomidine and midazolam sedation in mechanically ventilated patients. Time at the targeted sedation level was similar, while delirium prevalence during treatment was lower with dexmedetomidine, 54% versus 76.6% among the analysed patients. Extubation occurred earlier, although bradycardia was more frequent. These findings concern both the disadvantages of one sedation strategy and the benefits and harms of its replacement. They do not identify the independent effect of eliminating an intraoperative dose while leaving most other anaesthesia unchanged.³

DEXACET examined a different modifiable component of cardiac surgical care. In a small factorial trial involving 120 patients completing follow-up, scheduled intravenous paracetamol or placebo was combined with postoperative propofol or dexmedetomidine. Delirium occurred in 10% with paracetamol and 28% with placebo.

The dexmedetomidine–propofol comparison did not show a clear difference. This trial suggests that analgesic strategy may influence delirium, but its small size and distinct intervention prevent it from establishing a universal medication pathway. It reinforces the need to evaluate the whole perioperative context rather than attribute delirium to one drug class alone.⁶

PremedX studied preoperative lorazepam in 1,062 adults younger than 70 years undergoing elective surgery. Cardiac surgery was excluded. Lorazepam did not improve the next-day patient-experience score compared with placebo or no premedication, and was associated with slower extubation and less frequent early cognitive recovery. These outcomes are not equivalent to postoperative delirium, and the population differs from B-Free. The study is relevant chiefly because it separates a presumed benefit of routine anxiolytic premedication from evidence demonstrating that benefit.⁷

For awareness, BAG-RECALL offers a methodological contrast. It randomised 6,041 high-risk surgical patients to protocols based on bispectral-index monitoring or end-tidal anaesthetic concentration, with education and alert systems. Postoperative interviews identified definite awareness in seven of 2,861 versus two of 2,852 interviewed patients; superiority of the bispectral-index strategy was not established. This was not a benzodiazepine trial. It demonstrates that structured assessment can identify rare recall events and that awareness prevention depends on a broader anaesthetic strategy than the presence or absence of one amnestic drug.⁸

B-Free adds the largest directly relevant randomised evaluation of an intraoperative benzodiazepine policy in cardiac surgery. Its result does not overturn evidence about prolonged intensive care sedation, and earlier smaller trials do not convert its primary analysis into a positive finding. By the end of 2025, a restrictive institutional default was a reasonable option, but a proven clinically important reduction in delirium had not been established. Any policy should preserve justified individual exceptions, maintain adequate anaesthesia and use reliable postoperative assessment rather than assume that drug omission alone prevents acute brain dysfunction.

Does restricting intraoperative benzodiazepines prevent delirium after cardiac surgery?

B-Free did not establish a reduction in 72-hour delirium, although a modest benefit remains possible. Restriction is a reasonable policy option, but post hoc analyses and passive awareness reporting do not prove superiority or equivalent safety.

References

1. Spence J, Devereaux PJ, Lee SF, D'Aragon F, Avidan MS, Whitlock RP, et al; B-Free Investigators and the Canadian Perioperative Anesthesia Clinical Trials Group. Benzodiazepine-Free Cardiac Anesthesia for Reduction of Postoperative Delirium: A Cluster Randomized Crossover Trial. *JAMA Surg.* 2025;160(3):286-294.
2. Pandharipande PP, Pun BT, Herr DL, Maze M, Girard TD, Miller RR, et al. Effect of sedation with dexmedetomidine vs lorazepam on acute brain dysfunction in mechanically ventilated patients: the MENDS randomized controlled trial. *JAMA.* 2007;298(22):2644-2653.
3. Riker RR, Shehabi Y, Bokesch PM, Ceraso D, Wisemandle W, Koura F, et al. Dexmedetomidine vs midazolam for sedation of critically ill patients: a randomized trial. *JAMA.* 2009;301(5):489-499.
4. Spence J, et al; B-Free Investigators. Supplemental online content to the B-Free cluster randomized crossover trial. *JAMA Surg.* 2025.
5. Kleiman AM, Calgi MP, McNeil JS. Benzodiazepines Not Main Suspect in Cardiac Surgery Delirium. *JAMA Surg.* 2025;160(3):294-295.
6. Subramaniam B, Shankar P, Shaefi S, Mueller A, O'Gara B, Banner-Goodspeed V, et al. Effect of Intravenous Acetaminophen vs Placebo Combined With Propofol or Dexmedetomidine on Postoperative Delirium Among Older Patients Following Cardiac Surgery: The DEXACET Randomized Clinical Trial. *JAMA.* 2019;321(7):686-696.
7. Maurice-Szamburski A, Auquier P, Viarre-Oreal V, Cuvillon P, Carles M, Ripart J, et al. Effect of sedative premedication on patient experience after general anesthesia: a randomized clinical trial. *JAMA.* 2015;313(9):916-925.
8. Avidan MS, Jacobsohn E, Glick D, Burnside BA, Zhang L, Villafranca A, et al. Prevention of intraoperative awareness in a high-risk surgical population. *N Engl J Med.* 2011;365(7):591-600.

MIND

Arthur AS, Jahromi BS, Saphier PS, Nickele CM, Ryan RW, Vajkoczy P, et al; MIND Study Investigators and Collaborators. Minimally Invasive Surgery vs Medical Management Alone for Intracerebral Hemorrhage: The MIND Randomized Clinical Trial. JAMA Neurol. 2025;82(11):1113–1121.

Introduction

Evacuating an intracerebral haematoma has an intuitive appeal. Removing blood may reduce mass effect and exposure of surrounding brain to blood products, but an operation can also injure tissue, provoke further bleeding or introduce complications. The clinical value of evacuation therefore depends on more than the amount of blood removed. Location, timing, initial injury, surgical access and subsequent recovery all influence whether a technically successful procedure produces an outcome that matters to the patient.¹

Large trials of conventional evacuation did not establish an overall functional benefit, encouraging the development of approaches intended to reduce disruption of healthy brain. Catheter-based evacuation with thrombolysis was evaluated in MISTIE III, without a significant improvement in its primary functional endpoint.² More recently, ENRICH demonstrated better utility-weighted functional outcomes after early minimally invasive evacuation, with the benefit principally attributable to patients with lobar haemorrhage.³

These procedures use different access routes, devices, anatomical selection and time windows. A positive result with one strategy does not automatically validate another device or justify extending treatment to deep haemorrhages.

MIND compared endoscopically guided evacuation using the Artemis Neuro Evacuation Device with medical management alone. Most participants had deep haemorrhages, and intervention was permitted up to 72 hours after onset. The trial's substantial reduction in haematoma volume, neutral late functional outcome and favourable early exploratory result make it an important test of the relationship between surgical performance and durable benefit.¹

Synopsis

MIND was an open-label, multicentre randomised trial conducted at 32 institutions in the United States, Canada, Germany and Austria. Recruitment ran from February 2018 to August 2023. Adults aged 18–80 years were eligible with a spontaneous supratentorial intracerebral haemorrhage of 20–80 mL, premorbid modified Rankin Scale score of 0–1, National Institutes of Health Stroke Scale score of at least 6 and Glasgow Coma Scale score of 5–15. Initial imaging had to occur within 24 hours of symptom onset. Written consent was obtained from the patient or an authorised representative.¹

The published report describes changes during recruitment. A requirement for a stability scan at least six hours after initial imaging was removed after the first two years, during which 73 participants had been enrolled. Revised eligibility also excluded primary thalamic haemorrhage and specified additional clinical exclusions, including severe active infection, kidney failure and direct factor Xa inhibitor use. These changes are reported here as described in the principal publication, without an independent protocol audit.

An interactive web system allocated participants 2:1 to minimally invasive surgery plus medical management or medical management alone. A minimisation algorithm balanced allocation by haemorrhage location and Hemphill intracerebral haemorrhage score. Medical treatment followed contemporary guideline-based practice as determined by the treating team. Surgery was performed under general anaesthesia through a burr hole and endoscopic sheath, using image guidance to select a trajectory intended to minimise damage to uninvolved tissue. The Artemis device evacuated the clot under direct neuroendoscopic visualisation. Surgery was to occur within 72 hours of onset.

The primary efficacy endpoint was the ordinal modified Rankin Scale at 180 days. Although the scale ranges from 0, no symptoms, to 6, death, scores of 5 and 6 were combined for the analyses. A certified assessor unaware of allocation conducted the 180-day assessment. The primary safety endpoint was death within 30 days. Secondary outcomes included utility-weighted disability, dichotomised disability, procedure duration and hospital and intensive care stay. Earlier functional assessments and analyses by haemorrhage location were exploratory.

The primary efficacy analysis retained all randomised participants and used proportional-odds logistic regression. An odds ratio above 1 favoured surgery. The report specified two-sided 96% confidence intervals and one-sided P values, with an efficacy threshold of 0.02 in the adaptive design. Missing or out-of-window functional assessments were imputed using a longitudinal model. Secondary and exploratory efficacy analyses used the per-protocol population, while safety analyses classified

patients by treatment received. Their 95% confidence intervals were not adjusted for multiple comparisons.

Recruitment stopped before the originally intended completion. The scheduled interim analysis after 200 participants did not meet the predefined stopping criteria. Subsequently, announcement of the positive ENRICH results affected willingness to randomise patients with lobar haemorrhage. Lobar recruitment ceased, and a separate feasibility analysis indicated a low probability of demonstrating a difference if further recruitment comprised only deep haemorrhages. The investigators and sponsor decided to stop enrolment in October 2023 while continuing follow-up.

Of 4,066 screened patients, 236 were randomised: 154 to surgery and 82 to medical treatment. Median age was 60 years and median baseline NIHSS score 18. Deep haemorrhages accounted for 164 participants, 69.5%, predominantly involving the basal ganglia; 72 had primarily lobar haemorrhages. Investigator-estimated median haematoma volume was 39 mL, and 40.7% had intraventricular extension. The per-protocol population comprised 144 surgical and 78 medical participants. The as-treated safety population comprised 152 and 82, reflecting two crossovers and two withdrawals before surgery.

Among per-protocol surgical participants, median onset-to-procedure time was 27.5 hours. Of 142 with timing data, 55, 38.7%, underwent surgery within 24 hours and 15, 10.6%, within 12 hours. Median haematoma volume reduction was 80.7%, leaving a median residual volume of 6.3 mL. Table 2 reports a residual volume below 15 mL in 114 of 144 surgical participants, 79.2%. Median evacuation procedure duration, measured from sheath placement to removal, was 33.5 minutes. Conversion to craniotomy occurred in two participants, and significant procedural bleeding requiring cauterisation in nine.^{1,4}

The primary analysis did not demonstrate improved 180-day disability: odds ratio 1.03 (96% CI 0.62–1.72; one-sided $P=0.45$). Adjustment for the balancing strata gave an odds ratio of 1.10 (96% CI 0.66–1.85). There were 184 participants with in-window 180-day data; 52 scores were imputed, including 32 assessed outside the specified window and 20 not assessed. Per-protocol analyses were also inconclusive.

At 180 days, mean utility-weighted Rankin scores were 0.41 with surgery and 0.38 with medical management, with an adjusted difference of 0.04 (95% CI –0.04 to 0.12). In the per-protocol population, scores of 0–3 occurred in 36.8% versus 37.2%, and scores of 0–2 in 16.7% in each group. Intensive care and hospital stay showed no clear difference. An

exploratory adjusted per-protocol analysis favoured surgery at 30 days, with an odds ratio of 4.23 (95% CI 2.36–7.57), but the difference was not sustained at 90 or 180 days.

In the as-treated population, 30-day mortality was 11 of 152, 7.2%, after surgery versus eight of 82, 9.8%, with medical treatment: absolute difference –2.5 percentage points (95% CI –11.7 to 4.8). At 180 days, mortality was 13.2% versus 18.3%, also inconclusive. Serious adverse events by 180 days affected 52.6% versus 68.3%, a difference of –15.7 percentage points (95% CI –28.1 to –1.1). Penumbra funded the trial and participated in its design, conduct, data management, analyses and reporting. Independent committees oversaw safety and events, an independent imaging reviewer assessed scans, and Berry Consultants performed the primary efficacy analysis.

Critique

MIND separates technical success from clinical efficacy. The operation removed most of the haematoma in many treated participants, yet the primary estimate for late disability was close to no effect. The result therefore cannot be dismissed simply as an inability to achieve substantial evacuation. It also cannot prove that removing blood is intrinsically ineffective: the confidence interval remains compatible with meaningful benefit or harm, and the particular combination of device, timing and case mix matters. What it establishes is that this trial did not demonstrate durable functional benefit from its tested strategy.

Early termination is a major source of uncertainty. The scheduled interim analysis had supported continuation, whereas the later stop followed an external change in equipoise and a feasibility assessment under a proposed change in recruitment. This was not the same as crossing a prespecified futility boundary for the original trial population. The remaining sample consequently provides less precision, particularly within the 72-patient lobar subgroup. It would be inappropriate either to treat the trial as a definitive equivalence study or to presume that completion would have made the result positive.

ENRICH studied a different operation within an earlier window. Halting further lobar recruitment limited MIND's ability to establish whether its technique shared that benefit. The external result cannot substitute for this missing randomised evidence.

Timing is a plausible explanation for differing effects, but it remains an explanation rather than a finding of MIND. The median operation occurred after 24 hours, and only about one tenth of surgical participants were treated within 12 hours. Blood removal after established injury might offer less opportunity to alter recovery than earlier removal. Conversely, very early surgery may carry different bleeding risks. Participants

were not randomised between early and late evacuation, so comparisons by achieved surgical timing would be confounded by clinical stability, transfer, anatomy and logistics.

Haemorrhage location similarly requires restraint. About 70% of MIND participants had deep bleeds, the reverse of the approximate distribution in ENRICH. Damage to deep structures may be less reversible, and access may have different consequences from evacuation of a lobar clot. However, MIND's exploratory deep and lobar estimates were both imprecise and did not establish a location-specific benefit. A small subgroup without a clear effect cannot refute ENRICH's lobar result, while a favourable direction within that subgroup cannot establish that Artemis is effective there.

The 30-day disability signal deserves attention because its magnitude was substantial, but its methodological status differs from the primary endpoint. Earlier assessments were not blinded, the analysis used a per-protocol population and the confidence intervals were not adjusted for multiple exploratory comparisons. Knowledge that a patient had undergone surgery could influence judgement of functional status. Excluding participants with treatment or eligibility deviations also weakens the protection afforded by randomisation. These features do not prove that the signal is spurious, but they prevent it from carrying the same evidential weight as the masked 180-day assessment.

Its clinical meaning is also narrower than an early return to independence. The authors attribute the observed shift mainly to movement from Rankin score 5 to 4: from being bedridden and requiring constant care to still being unable to walk or attend to bodily needs without assistance. That change could matter to patients and families, but it remains substantial dependence. The supplementary distributions show later improvement in the medical group and convergence of outcomes.⁴ The evidence therefore suggests a possible earlier change in the degree of severe disability, without establishing lasting independence or a sustained quality-of-life advantage.

Combining scores of 5 and 6 further affects interpretation. The ordinal model does not distinguish death from survival with the most severe disability at its worst level. The utility-weighted analysis similarly assigns both states a weight of zero. These are explicit analytical choices, not evidence that patients regard the states identically. Mortality therefore needs separate consideration, and the survival estimates were too imprecise to demonstrate a mortality benefit or exclude a clinically important adverse difference.

Missing assessments introduce a further dependency on modelling. Fifty-two of 236 primary scores, approximately 22%, were imputed, although most of these represented out-of-window rather than completely absent assessments. Those circumstances differ

from losing all information on a participant, but imputation still assumes that the available longitudinal information adequately represents the missing time point. The published account does not provide enough detail to independently assess every feature of that model. The primary result is not invalidated by this alone; confidence in a precise null conclusion is nevertheless limited.

Safety cannot be summarised as “the procedure was safe” simply because mortality did not increase significantly. The 30-day interval includes both a substantial reduction and an increase of several percentage points in death. Significant procedural bleeding requiring cauterisation occurred in 6.3% of per-protocol surgical patients, and a small number required conversion to craniotomy. These are meaningful procedural observations even when successful evacuation is achieved. The lower overall serious-adverse-event rate is encouraging, but arose from an as-treated, nonprimary comparison without multiplicity adjustment and does not establish the mechanism of any advantage.

The trial incorporated protections against biased assessment, including concealed web allocation, balancing of key prognostic variables, independent imaging review and masked primary functional assessment. Those strengths are particularly valuable in an operation that cannot practically be concealed from the treating team. Sponsor participation extended well beyond supplying the device, encompassing analysis and manuscript development; independent primary analysis and endpoint oversight therefore matter. Funding and participation should be transparent without being used as a substitute for examining the actual design and results.

Applicability is limited to selected adults who were largely independent before their haemorrhage and had neither very small nor exceptionally large clots. Only 236 of more than 4,000 screened patients were randomised. Almost all deep bleeds involved the basal ganglia, with primary thalamic haemorrhage excluded after amendment. The trial cannot establish the best rescue procedure for impending herniation, the management of infratentorial haemorrhage or outcomes in patients older than 80 years. It also cannot show whether another minimally invasive route or decompression without clot evacuation would produce the same result.

MIND therefore supports continued clinical investigation rather than routine expansion of this particular evacuation strategy. Future work needs to preserve randomised comparisons while addressing timing, anatomical selection and outcomes important to patients. Achieving a low residual clot volume is a useful measure of procedural delivery; it remains insufficient as the sole justification for expecting functional recovery.

Body of Evidence

The 2005 STICH trial randomised 1,033 patients to early surgery or initial conservative management, with later evacuation allowed when necessary. At six months, favourable outcomes occurred in 26% versus 24% of assessed patients, without a clear overall benefit.⁵ This established the limitations of assuming that conventional evacuation should improve recovery across a broad supratentorial population. It did not test every subsequent minimally invasive technique.

STICH II subsequently focused on conscious patients with superficial lobar haemorrhage and no intraventricular extension. Among 601 randomised patients, the primary unfavourable outcome occurred in 59% of analysed early-surgery participants and 62% of controls, an absolute difference of 3.7 percentage points (95% CI -4.3 to 11.6).⁶ Anatomical restriction did not establish a significant primary functional benefit, although the estimate left uncertainty about smaller effects. Both STICH trials used prognosis-based Glasgow outcome measures, which should not be equated directly with the ordinal Rankin analysis in MIND.

MISTIE III evaluated a substantially different approach: image-guided catheter evacuation followed by local alteplase, targeting a residual clot of 15 mL or less. It randomised 506 adults with supratentorial haemorrhage of at least 30 mL; 499 entered the modified intention-to-treat efficacy analysis. The adjusted proportion estimated to achieve Rankin 0–3 at one year was 45% with MISTIE and 41% with medical care, an absolute difference of 4 percentage points (95% CI -4 to 12).² This neutral primary result reinforced the need to distinguish successful removal and possible survival signals from demonstrated functional benefit.

Improved medical treatment also changes the context in which surgery is evaluated. INTERACT3 used a stepped-wedge cluster design to implement a bundle addressing blood pressure, glucose, temperature and abnormal anticoagulation. Its modified intention-to-treat population comprised 7,036 patients at 121 hospitals. The care bundle improved the distribution of six-month Rankin scores, with a common odds ratio for poorer outcome of 0.86 (95% CI 0.76–0.97).⁷ This was not a surgical trial and does not identify why MIND was neutral. It demonstrates that the medical comparator is active treatment and that organised physiological care can influence recovery.

ENRICH supplied the most important positive evacuation evidence before MIND's publication. Three hundred patients with lobar or anterior basal ganglia haemorrhage of 30–80 mL were randomised to early minimally invasive removal plus medical management or medical management alone. Surgery was deliverable within 24 hours. After an adaptive rule was triggered, further enrolment was restricted to lobar

haemorrhage. Mean utility-weighted Rankin score at 180 days was 0.458 with surgery and 0.374 with medical treatment: difference 0.084 (95% Bayesian credible interval 0.005–0.163), meeting the specified probability threshold for superiority.³

The ENRICH effect was concentrated in lobar haemorrhage, where the estimated difference was 0.127; the anterior basal ganglia estimate was –0.013. Thirty-day mortality was 9.3% versus 18.0%. These findings support a defined early operative approach, particularly in lobar disease. Compared with MIND, the study used different equipment and access, treated earlier, enrolled proportionally more lobar bleeds and required a larger minimum clot volume. Cross-trial comparisons cannot determine which difference explains the divergent results, nor establish that one device is superior to the other.

The 2024 SWITCH trial addressed another surgical question in severe deep haemorrhage: decompressive craniectomy plus medical treatment versus medical treatment alone. Recruitment ended early because of funding limitations. Of 201 randomised participants, 197 provided delayed consent. Rankin 5–6 at 180 days occurred in approximately 44% versus 58%, with an adjusted risk ratio of 0.77 (95% CI 0.59–1.01).⁸ This suggested possible benefit but remained inconclusive, with severe disability common. Decompression and clot evacuation are distinct interventions, and SWITCH cannot be used to supply a missing positive result for Artemis.

Taken together, the evidence available by the end of 2025 favoured increasingly specific questions about location, timing and operative method. ENRICH narrowed uncertainty for an early approach to selected lobar haemorrhage. MIND showed that substantial endoscopic evacuation, predominantly in deep haemorrhage and often after 24 hours, did not automatically yield durable functional improvement. The field has moved beyond asking whether a clot can be removed; the unresolved issue is when removal produces a patient-valued advantage over effective medical care.

Should Artemis-guided evacuation within 72 hours be used routinely for moderate-to-large supratentorial intracerebral haemorrhage?

Not on the basis of MIND: substantial clot removal did not demonstrate better 180-day function. Its result should be distinguished from evidence for earlier surgery using a different approach in selected lobar haemorrhage.

References

1. Arthur AS, Jahromi BS, Saphier PS, Nickle CM, Ryan RW, Vajkoczy P, et al; MIND Study Investigators and Collaborators. Minimally Invasive Surgery vs Medical Management Alone for Intracerebral Hemorrhage: The MIND Randomized Clinical Trial. *JAMA Neurol.* 2025;82(11):1113-1121.
2. Hanley DF, Thompson RE, Rosenblum M, Yenokyan G, Lane K, McBee N, et al. Efficacy and safety of minimally invasive surgery with thrombolysis in intracerebral haemorrhage evacuation (MISTIE III): a randomised, controlled, open-label, blinded endpoint phase 3 trial. *Lancet.* 2019;393(10175):1021-1032.
3. Pradilla G, Ratcliff JJ, Hall AJ, Saville BR, Allen JW, Paulon G, et al. Trial of Early Minimally Invasive Removal of Intracerebral Hemorrhage. *N Engl J Med.* 2024;390(14):1277-1289.
4. Arthur AS, Jahromi BS, Saphier PS, et al; MIND Study Investigators and Collaborators. Supplement 2 to: Minimally Invasive Surgery vs Medical Management Alone for Intracerebral Hemorrhage: The MIND Randomized Clinical Trial. *JAMA Neurol.* 2025;82(11):1113–1121.
5. Mendelow AD, Gregson BA, Fernandes HM, Murray GD, Teasdale GM, Hope DT, et al. Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the International Surgical Trial in Intracerebral Haemorrhage (STICH): a randomised trial. *Lancet.* 2005;365(9457):387-397.
6. Mendelow AD, Gregson BA, Rowan EN, Murray GD, Gholkar A, Mitchell PM. Early surgery versus initial conservative treatment in patients with spontaneous supratentorial lobar intracerebral haematomas (STICH II): a randomised trial. *Lancet.* 2013;382(9890):397-408.
7. Ma L, Hu X, Song L, Chen X, Ouyang M, Billot L, et al. The third Intensive Care Bundle with Blood Pressure Reduction in Acute Cerebral Haemorrhage Trial (INTERACT3): an international, stepped wedge cluster randomised controlled trial. *Lancet.* 2023;402(10395):27-40.
8. Beck J, Fung C, Strbian D, Bütikofer L, Z'Graggen WJ, Lang MF, et al. Decompressive craniectomy plus best medical treatment versus best medical treatment alone for spontaneous severe deep supratentorial intracerebral haemorrhage: a randomised controlled clinical trial. *Lancet.* 2024;403(10442):2395-2404.

POST-TNK

Huang J, Yang J, Liu C, Li L, Yang D, Guo C, et al. Intra-Arterial Tenecteplase Following Endovascular Reperfusion for Large Vessel Occlusion Acute Ischemic Stroke: The POST-TNK Randomized Clinical Trial. JAMA. 2025;333(7):579–588.

Introduction

A technically successful thrombectomy does not guarantee recovery after ischaemic stroke. Some tissue may already be irreversibly injured before the artery opens. Other regions may remain inadequately perfused because of distal emboli, microvascular obstruction or disturbances of tissue perfusion that are not resolved by removing the proximal clot. An angiogram showing near-complete or complete reperfusion therefore leaves an important therapeutic question: can a further intervention improve tissue recovery without introducing unacceptable bleeding?

Adjunctive intra-arterial thrombolysis is intended to dissolve residual thrombus beyond the reach of thrombectomy devices. It can be delivered through an existing catheter immediately after the main procedure. This differs from rescue thrombolysis for a visibly persistent large arterial occlusion and from intravenous thrombolysis given before thrombectomy. Evidence for one timing or route does not establish benefit for another.

The small CHOICE trial provided an encouraging signal for intra-arterial alteplase after successful thrombectomy, with more patients reaching an excellent functional outcome at 90 days. Its early termination and imprecise estimate required confirmation. Tenecteplase was an attractive alternative because of its fibrin specificity and prolonged activity, but the appropriate intra-arterial dose and clinical effect remained uncertain.¹

POST-TNK tested a low dose of intra-arterial tenecteplase after near-complete or complete anterior-circulation reperfusion, specifically in patients who had not received intravenous thrombolysis. Its findings are most relevant to this defined procedural situation. They do not resolve every possible use of a thrombolytic during endovascular stroke treatment.²

Synopsis

POST-TNK was an investigator-initiated, multicentre, randomised trial at 34 comprehensive stroke centres in China. Treatment was open label, with blinded functional outcome assessment and independent adjudication. Recruitment occurred from October 2022 to March 2024, with final follow-up in June 2024. Public research funding supported the study, and the manufacturer supplied tenecteplase. The funders and suppliers had no reported role in design, conduct, analysis or publication decisions. The principal report appeared online on 13 January 2025.²

Eligible patients were adults with an intracranial internal carotid, first-segment middle cerebral artery or second-segment middle cerebral artery occlusion, treated within 24 hours of last known well. Baseline National Institutes of Health Stroke Scale (NIHSS) score had to be 25 or lower. Imaging required an Alberta Stroke Program Early CT Score (ASPECTS) of at least 6 within six hours, or a higher score or specified clinical–imaging mismatch criteria in the later window. Patients were randomised only after thrombectomy achieved an expanded Thrombolysis in Cerebral Infarction (eTICI) grade of 2c or 3, representing approximately 90–99% or complete reperfusion.^{2,3}

Important exclusions were intravenous thrombolysis before thrombectomy, prestroke modified Rankin Scale (mRS) score of 2 or higher, intracranial haemorrhage, procedural vascular injury, a procedure lasting more than 90 minutes and more than three stent-retriever passes. Other exclusions addressed major bleeding risk, severe renal impairment and complete clinical recovery after rapid reperfusion. These restrictions selected patients with little preceding disability and relatively uncomplicated successful procedures. A few eligibility violations were identified subsequently.^{2,3}

A web-based system assigned patients equally using varying permuted block sizes. Intervention patients received tenecteplase 0.0625 mg/kg, to a maximum of 6.25 mg, through a distal access catheter or microcatheter positioned proximal to the initially occluded artery. It was infused over 10–15 minutes. This was not the usual intravenous bolus regimen. In control patients the procedure ended without adjunctive intra-arterial thrombolysis; there was no placebo infusion.²

The primary efficacy endpoint was freedom from significant disability at 90 days, defined as mRS 0 or 1. The mRS ranges from 0, no symptoms, to 6, death. Assessment used centrally reviewed video or audio recordings where available, with trained local blinded assessors when recordings were unavailable. Secondary efficacy endpoints included functional independence, mRS 0–2; the overall mRS distribution; neurological change at days 5–7 or earlier discharge; and health-related quality of life.²

The primary safety outcomes were all-cause mortality through 90 days and symptomatic intracranial haemorrhage within 48 hours. The modified Heidelberg definition required haemorrhage on imaging with neurological deterioration of at least four total NIHSS points or at least two points in an NIHSS category. Any radiological intracranial haemorrhage, bleeding classifications and adverse events were also assessed. Safety and secondary analyses were exploratory, without multiplicity adjustment.²

The sample-size calculation assumed 32.8% of control patients would achieve mRS 0–1 and sought an absolute increase of 13 percentage points, smaller than the initial CHOICE estimate but still substantial. With 83% power and a two-sided significance level of 0.05, the target after allowance for attrition was 498 patients. Recruitment was extended when enrolment was brisk and additional study drug was available, with monitoring-board approval. The report does not describe stopping because an efficacy boundary had been crossed.²

Primary binary outcomes were analysed using modified Poisson regression, adjusting for age, baseline NIHSS, baseline ASPECTS, occlusion site and time from last known well to randomisation. The ordinal mRS analysis used a generalised odds ratio, which compares the odds of a better versus worse outcome across patient pairs, rather than a conventional proportional-odds model. Other non-normal outcomes used win-ratio methods, and mortality used Cox regression. Risk differences were calculated in post hoc analyses.²

Of 661 patients assessed, 541 were randomised. One tenecteplase-assigned patient immediately withdrew consent, leaving 269 intervention and 271 control participants. One control participant lacked a 90-day assessment, so the primary analysis included 539 patients, 269 versus 270. Per-protocol analyses retained 252 versus 243 after excluding major deviations. Sensitivity analyses considered missing outcomes, covariate adjustment and centre effects.^{2,3}

Median age was 69 years and median NIHSS score 15 in both groups. Median ASPECTS was 8. Approximately half the strokes were classified as cardioembolic and two-fifths as large-artery atherosclerotic. The M1 segment was the most frequent occlusion site. Approximately 62% had complete eTICI 3 reperfusion and 38% near-complete eTICI 2c reperfusion. Median time from last known well to randomisation was 500 versus 490 minutes, with tenecteplase starting at a median of 504 minutes in the intervention group.²

At 90 days, mRS 0–1 occurred in 132/269 patients receiving tenecteplase and 119/270 controls, 49.1% versus 44.1%. The adjusted risk ratio was 1.15 (95% confidence interval

[CI] 0.97–1.36; P=0.11). The unadjusted absolute difference was 5.00 percentage points (95% CI –3.42 to 13.41). The primary analysis therefore did not establish superiority. Per-protocol and missing-data analyses were similar.^{2,3}

Functional independence occurred in 165/269 versus 159/270, adjusted risk ratio 1.06 (95% CI 0.93–1.21). The generalised odds ratio for the full mRS distribution was 1.22 (95% CI 0.96–1.55). Neurological change and quality-of-life analyses did not establish benefit. Deaths were 43/269 versus 52/270, 16.0% versus 19.3%, with an adjusted hazard ratio of 0.75 (95% CI 0.50–1.13).²

Symptomatic intracranial haemorrhage occurred in 17/268 versus 12/271 patients with the relevant data, 6.3% versus 4.4%; adjusted risk ratio 1.43 (95% CI 0.68–2.99). Any radiological intracranial haemorrhage occurred in 98/268 versus 74/271, 36.6% versus 27.3%; adjusted risk ratio 1.33 (95% CI 1.04–1.69). The unadjusted absolute increase was 9.26 percentage points (95% CI 1.43–17.10). No convincing treatment interaction was identified across the examined subgroups.²

Critique

POST-TNK is a substantial randomised evaluation of a well-defined adjunctive strategy. It enrolled substantially more patients than CHOICE, achieved almost complete 90-day follow-up and assessed a clinically important outcome with blinded raters. The narrow reperfusion criterion improves clarity: the question was whether additional thrombolysis helps after most or all of the angiographic territory has already reperfused. Treatment allocation occurred at that decision point, which avoids comparing patients with fundamentally different thrombectomy results.

The result is neutral, but not an equivalence finding. The primary confidence interval allows a clinically important benefit as well as little effect. The observed five-point absolute difference was much smaller than the 13-point improvement used for planning, and the control outcome rate was higher than anticipated. The trial was consequently not designed to settle whether a modest improvement exists. However, plausible benefit remaining within a confidence interval is not evidence sufficient for routine treatment, particularly when bleeding remains a concern.

A dichotomous mRS 0–1 endpoint is clinically understandable but discards information. Moving from severe dependency to walking with help is important, yet does not count as primary success. Conversely, a shift across the boundary between slight disability and no significant disability can determine the primary result. POST-TNK's full-scale and mRS

0–2 analyses were directionally compatible with some benefit but did not establish it. They provide useful context without converting the primary result into a positive trial.

The generalised odds ratio should be named accurately. It does not have the same interpretation as a common odds ratio from proportional-odds regression, even though both summarise an ordinal scale and values above one can favour treatment. Mixing these measures in comparisons or pooled analyses can create a misleading impression of identical estimands. The generalised odds ratio, the binary risk ratio and the mortality hazard ratio each address a different aspect of outcome. Their numerical magnitudes are not directly interchangeable.

The distinction between macrovascular and tissue reperfusion remains unresolved. An eTICI 2c–3 result does not prove that all threatened tissue receives adequate flow. Equally, poor recovery after such a result does not prove that persistent fibrin thrombi are the explanation. Infarct already present, oedema, endothelial injury and other microvascular processes may be unresponsive to fibrinolysis. POST-TNK did not require repeat angiography or perfusion imaging after tenecteplase, so it cannot show that the drug improved the proposed biological target even when the clinical outcome was favourable.²

The population also had relatively little residual angiographic abnormality to treat. Approximately three-fifths had complete eTICI 3 reperfusion. A therapy directed at distal clot might have a different effect in eTICI 2b patients, who were intentionally excluded. Conversely, including broader reperfusion grades in another trial may introduce a mixture of visible distal-artery treatment and true microvascular rescue. Those are related but not identical mechanisms. The absence of a clear 2c-versus-3 interaction here does not establish that reperfusion grade is irrelevant; subgroup estimates were imprecise.

The dose was one-quarter of the commonly studied intravenous tenecteplase dose of 0.25 mg/kg, delivered as a local infusion rather than a bolus. Choosing a lower dose after reperfusion is reasonable as a risk-management hypothesis, but POST-TNK was not a dose-finding study. Its result cannot demonstrate that a higher dose would succeed, nor that the chosen dose was biologically inactive. Any proposal to escalate exposure must account for the observed increase in radiological haemorrhage.^{2,4}

Bleeding deserves more attention than a statement that there was no significant difference in symptomatic haemorrhage. Only 29 symptomatic events occurred, and the confidence interval for their relative risk extended from a reduction to almost a tripling. This is insufficient to establish safety equivalence. The increase in all radiological

intracranial haemorrhage is a secondary, multiplicity-unadjusted result, so it should not be overstated as definitive proof of clinical harm. Nonetheless, it is biologically plausible and argues against describing the treatment as demonstrably harmless.

Nor should all radiological bleeding be assumed to be inconsequential. Its immediate clinical meaning varies, but haemorrhage can influence subsequent antithrombotic treatment, monitoring and recovery. The study did not establish that the excess imaging abnormalities had no later consequences. Similarly, the numerical mortality difference favouring tenecteplase does not establish a survival benefit. A balanced interpretation retains uncertainty around both benefit and important harm rather than using non-significant safety tests as reassurance.

Open-label delivery is another limitation. Blinded central outcome assessment reduces detection bias, but treating teams knew whether additional thrombolysis had been given. This could influence blood-pressure management, antithrombotic timing, imaging, monitoring or later care. The comparator also ended the procedure while intervention patients remained for a 10–15-minute infusion. A placebo-controlled design would have isolated the pharmacological effect more cleanly. The report does not provide enough detail to establish that all consequential postprocedural care was identical.

Generalisability depends on clinical features, not simply the country of recruitment. The high prevalence of intracranial atherosclerotic mechanisms may matter, as may the exclusion of prior intravenous thrombolysis, severe preceding disability, large early infarction and difficult or prolonged procedures. The findings do not apply directly to patients receiving standard bridging thrombolysis or to those with substantial residual occlusion. Those groups should not be treated by extrapolating a favourable trend from this selected population.

The small amount of missing primary data is unlikely to explain the result: alternative assumptions and imputation barely changed the risk ratio. The per-protocol analysis was also neutral, but it should remain supportive because removing patients after allocation can introduce bias. Six patients had received intravenous thrombolysis despite the intended exclusion, among other deviations. Maintaining their outcomes in the main retained population better preserves the randomised comparison than retrospectively constructing a perfectly adherent subgroup.³

Finally, the parallel POST-UK trial cannot establish which thrombolytic is preferable. Similar eligibility and contemporaneous conduct make it relevant, but the agents were not compared within one randomised allocation. Differences in mortality or bleeding between separate trials may reflect chance, case mix or management. The accompanying

editorial highlights the value of a direct multiarm comparison. Its suggestion of a common thrombolytic effect is a hypothesis that must be weighed against differences in agent, dose, reperfusion and previous intravenous treatment.⁵

Body of Evidence

CHOICE randomised 121 patients after successful thrombectomy in seven Spanish centres; 113 were treated as allocated. Intra-arterial alteplase, 0.225 mg/kg to a maximum of 22.5 mg, was compared with placebo after eTICI 2b50–3 reperfusion. Excellent outcome occurred in 36/61 versus 21/52 patients, 59.0% versus 40.4%, with an adjusted absolute difference of 18.4 percentage points (95% CI 0.3–36.4). The trial stopped early because of pandemic-related recruitment and placebo-supply difficulties. The signal justified further trials, but its magnitude was uncertain and the broader disability analyses were less persuasive.¹

POST-UK tested 100 000 IU of intra-arterial urokinase after eTICI 2c–3 reperfusion, also excluding prior intravenous thrombolysis. Among 532 patients with 90-day outcomes, mRS 0–1 occurred in 120/266 versus 107/266, adjusted risk ratio 1.13 (95% CI 0.94–1.36). Symptomatic haemorrhage was approximately 4% in both groups. This independent randomised comparison gave a similarly imprecise, modestly favourable primary point estimate without establishing superiority. It supports continuing uncertainty about routine adjunctive treatment after near-complete reperfusion.⁶

ATTENTION-IA evaluated the same tenecteplase dose as POST-TNK after posterior-circulation thrombectomy, but allowed eTICI 2b50–3 reperfusion and used a brief injection. In 208 patients, mRS 0–1 occurred in 36/104 versus 27/104, adjusted risk ratio 1.36 (95% CI 0.92–2.02). Symptomatic intracranial haemorrhage was numerically more frequent with tenecteplase, with a wide confidence interval. Posterior-circulation anatomy and case severity differ from POST-TNK, but the trial similarly failed to establish clinical benefit or exclude an important bleeding risk.⁷

Later in 2025, ANGEL-TNK provided a positive primary result with a different regimen and population. It used intra-arterial tenecteplase 0.125 mg/kg after eTICI 2b–3 reperfusion in anterior-circulation strokes treated 4.5–24 hours after last known well. Among 255 patients with outcomes, mRS 0–1 occurred in 51/126 versus 34/129, with a reported adjusted relative risk of 1.44 (95% CI 1.06–1.95). None of its secondary efficacy analyses established benefit. The result raises a dose and selection hypothesis, but it is not a randomised demonstration that doubling the POST-TNK dose improves outcomes.⁸

PEARL subsequently randomised 324 patients after successful anterior-circulation thrombectomy to intra-arterial alteplase, 0.225 mg/kg to a maximum of 20 mg, or standard treatment. It allowed eTICI 2b50–3 reperfusion and previous intravenous thrombolysis. Excellent outcome occurred in 73/163 versus 48/159 patients, adjusted risk ratio 1.45 (95% CI 1.08–1.96). Mortality and any intracranial haemorrhage were numerically higher with alteplase, although neither difference was statistically clear. PEARL strengthens the efficacy signal for some adjunctive strategies while retaining uncertainty about the overall balance and optimal population.⁹

EXTEND-IA TNK addresses a separate question. In 202 patients treated intravenously before thrombectomy, tenecteplase 0.25 mg/kg improved early reperfusion compared with alteplase, 22% versus 10%, with a favourable secondary functional analysis. This supports tenecteplase's thrombolytic activity in a proximal occlusion before mechanical treatment. It cannot establish efficacy for a much lower intra-arterial dose after the artery has already been reopened.⁴

By the end of 2025, the evidence was therefore mixed rather than uniformly negative. Several trials suggested more excellent outcomes, but effects were not consistently reproduced across broader disability measures, agents, doses or reperfusion grades. POST-TNK specifically does not support routine low-dose intra-arterial tenecteplase after eTICI 2c–3 reperfusion in patients without prior intravenous thrombolysis. Its imprecision leaves room for further study, while the radiological bleeding signal makes empirical adoption less attractive.

Should low-dose intra-arterial tenecteplase be added after near-complete or complete thrombectomy reperfusion?

Not routinely on the basis of POST-TNK: functional benefit was not established and radiological intracranial bleeding increased. Other doses and patient selections remain separate questions requiring confirmatory evidence.

References

1. Renú A, Millán M, San Román L, Blasco J, Martí-Fàbregas J, Terceño M, et al. Effect of Intra-arterial Alteplase vs Placebo Following Successful Thrombectomy on Functional Outcomes in Patients With Large Vessel Occlusion Acute Ischemic Stroke: The CHOICE Randomized Clinical Trial. *JAMA*. 2022;327(9):826-835.

2. Huang J, Yang J, Liu C, Li L, Yang D, Guo C, et al. Intra-Arterial Tenecteplase Following Endovascular Reperfusion for Large Vessel Occlusion Acute Ischemic Stroke: The POST-TNK Randomized Clinical Trial. *JAMA*. 2025;333(7):579-588.
3. Huang J, Yang J, Liu C, Li L, Yang D, Guo C, et al. Supplementary material to: Intra-Arterial Tenecteplase Following Endovascular Reperfusion for Large Vessel Occlusion Acute Ischemic Stroke: The POST-TNK Randomized Clinical Trial. *JAMA*. 2025;333(7):579–588.
4. Campbell BCV, Mitchell PJ, Churilov L, Yassi N, Kleinig TJ, Dowling RJ, et al. Tenecteplase versus Alteplase before Thrombectomy for Ischemic Stroke. *N Engl J Med*. 2018;378(17):1573-1582.
5. Dippel DWJ, Khan CF, Schoon BA. Intra-Arterial Thrombolytics During Thrombectomy for Ischemic Stroke-End of the Story or a New Beginning?. *JAMA*. 2025;333(7):571-573.
6. Liu C, Guo C, Li F, Yu N, Huang J, Peng Z, et al. Intra-Arterial Urokinase After Endovascular Reperfusion for Acute Ischemic Stroke: The POST-UK Randomized Clinical Trial. *JAMA*. 2025;333(7):589-598.
7. Hu W, Tao C, Wang L, Chen Z, Li D, Chen W, et al. Intra-arterial tenecteplase after successful endovascular recanalisation in patients with acute posterior circulation arterial occlusion (ATTENTION-IA): multicentre randomised controlled trial. *BMJ*. 2025;388:e080489.
8. Miao Z, Luo G, Song L, Sun D, Chen W, Yao X, et al. Intra-arterial Tenecteplase for Acute Stroke After Successful Endovascular Therapy: The ANGEL-TNK Randomized Clinical Trial. *JAMA*. 2025;334(7):582-591.
9. Writing Committee for the PEARL Investigators. Intra-Arterial Alteplase After Successful Endovascular Reperfusion in Acute Stroke: The PEARL Randomized Clinical Trial. *JAMA*. 2025;334(19):1728–1739.

POST-UK

Liu C, Guo C, Li F, Yu N, Huang J, Peng Z, et al. Intra-Arterial Urokinase After Endovascular Reperfusion for Acute Ischemic Stroke: The POST-UK Randomized Clinical Trial. JAMA. 2025;333(7):589–598.

Introduction

After thrombectomy restores flow through an occluded cerebral artery, there is a natural temptation to optimise the result further. Small distal emboli may remain beyond the reach of a retrieval device, and angiographic reperfusion may not correspond to complete restoration of flow through threatened brain tissue. A locally delivered fibrinolytic could potentially dissolve residual clot and improve recovery. However, exposing recently ischaemic tissue to additional thrombolysis may also increase haemorrhage.

Urokinase has a long history in intra-arterial stroke treatment. Earlier studies evaluated it as a way to reopen an occluded artery, and observational series subsequently described its use when thrombectomy failed or left incomplete reperfusion. Neither situation is equivalent to routinely adding urokinase after near-complete or complete reperfusion. The amount and location of remaining thrombus, the condition of the tissue and the balance between benefit and bleeding may all differ.^{1,2}

The CHOICE trial renewed interest in adjunctive intra-arterial treatment after thrombectomy by suggesting that alteplase could increase excellent functional outcomes. It was small and stopped early, so confirmation was needed. Urokinase offered an alternative agent with practical relevance in settings where it was already available, but its efficacy could not be assumed from an alteplase result or from experience with rescue treatment.³

POST-UK therefore evaluated a fixed low dose of intra-arterial urokinase after successful anterior-circulation thrombectomy in patients who had not received intravenous thrombolysis. The trial asks whether adding a drug at the end of an apparently successful procedure improves survival without significant disability, rather than whether urokinase can recanalise a persistently occluded large artery.⁴

Synopsis

POST-UK was an investigator-initiated, multicentre, randomised, open-label trial with blinded endpoint assessment at 35 comprehensive stroke centres in China. Recruitment ran from November 2022 to March 2024, and final follow-up was completed in July 2024. Independent academic investigators led the study, with monitoring and adjudication committees. Chinese public research programmes funded the work; Wuhan Humanwell Pharmaceutical supplied the study drug. The sponsors and supplier had no reported role in study conduct, analysis or publication decisions.⁴

Adults were eligible after thrombectomy for occlusion of the intracranial internal carotid artery or the first or second segment of the middle cerebral artery. They had to be enrolled within 24 hours of last known well, have a baseline National Institutes of Health Stroke Scale (NIHSS) score of 25 or lower, and have been independent in usual activities before stroke, corresponding to a modified Rankin Scale (mRS) score below 2. Prior intravenous thrombolysis was an exclusion.^{4,5}

Imaging selection required a small-to-moderate ischaemic core: an Alberta Stroke Program Early CT Score (ASPECTS) of at least 6 within six hours, or at least 7 or specified clinical-imaging mismatch criteria between six and 24 hours. Randomisation followed near-complete or complete reperfusion, an expanded Thrombolysis in Cerebral Infarction (eTICI) grade of 2c or 3. This represents approximately 90–99% or complete reperfusion of the angiographic territory. Patients with more substantial residual perfusion deficits were not the intended population.^{4,5}

Other important exclusions included intracranial haemorrhage, vascular rupture or dissection during the procedure, procedure duration exceeding 90 minutes, more than three stent-retriever passes, major bleeding risk and severe renal impairment. These criteria reduced the inclusion of difficult procedures and patients at especially high haemorrhagic risk. Consent was obtained from patients or authorised representatives before allocation.^{4,5}

A web-based system randomised patients equally using varying block sizes. The intervention was a single 100 000 IU dose of urokinase, reconstituted in 10 mL of saline and infused over 10–15 minutes through a distal access catheter or microcatheter proximal to the originally occluded artery. Controls ended the procedure without intra-arterial thrombolysis. There was no placebo infusion. Thrombectomy techniques followed local practice, and concomitant medical therapy was expected to follow national stroke guidance.⁴

The primary efficacy endpoint was survival without significant disability at 90 days, defined as mRS 0 or 1. The scale ranges from 0, no symptoms, to 6, death. Certified blinded evaluators assessed recordings centrally, with local blinded assessment when recordings were unavailable. Secondary outcomes were functional independence, mRS 0–2; the complete mRS distribution; NIHSS change by days 5–7 or earlier discharge; and health-related quality of life at 90 days.⁴

The primary safety endpoints were all-cause mortality by 90 days and symptomatic intracranial haemorrhage within 48 hours. The modified Heidelberg definition linked an imaging-confirmed haemorrhage to at least a four-point deterioration in total NIHSS or a two-point deterioration in a subcategory. Any radiological intracranial haemorrhage, bleeding events and adverse events were also recorded. Secondary and safety analyses were exploratory, without correction for multiple comparisons.⁴

Planning assumed an excellent outcome rate of 32.8% in controls and a 13-percentage-point absolute improvement with urokinase. The target was 498 patients after allowing for attrition, providing 83% power at a two-sided significance level of 0.05. With brisk recruitment and additional drug availability, enrolment was extended with monitoring-board approval. The study was not terminated early for an observed treatment benefit.⁴

The main analysis used complete 90-day cases from the retained randomised population. Modified Poisson regression estimated binary-outcome risk ratios, adjusted for age, baseline NIHSS, baseline ASPECTS, occlusion site and time to randomisation. An analysis of the full mRS used a generalised odds ratio rather than a proportional-odds model. Neurological change and quality of life used win ratios, with inverse-probability weighting for adjusted comparisons. Mortality was analysed using Cox regression. Absolute risk differences were calculated post hoc.⁴

Of 756 patients assessed, 535 were randomised, 268 to urokinase and 267 to control. One intervention patient immediately withdrew consent, leaving 267 in each group. There was no treatment crossover. One patient per group lacked a 90-day assessment, leaving 266 per group in the primary analysis. Per-protocol analyses included 249 per group. Alternative missing-data assumptions, multiple imputation and a centre-adjusted model were examined.^{4,5}

Median age was 69 years in the urokinase group and 68 in controls. Median NIHSS was 15 and ASPECTS 8 in both groups. Approximately half the strokes were attributed to large-artery atherosclerosis and approximately two-fifths to cardioembolism. Most occlusions involved the M1 segment. Around two-thirds had complete eTICI 3

reperfusion, with most remaining patients at grade 2c. Median time from last known well to randomisation was 523 versus 524 minutes.⁴

At 90 days, mRS 0–1 occurred in 120/266 urokinase patients and 107/266 controls, 45.1% versus 40.2%. The adjusted risk ratio was 1.13 (95% confidence interval [CI] 0.94–1.36; $P=0.19$). The unadjusted absolute difference was 4.89 percentage points (95% CI –3.51 to 13.28). Per-protocol results were similar, with an adjusted risk ratio of 1.15 (95% CI 0.95–1.39), and missing-data sensitivity analyses did not materially change the interpretation.^{4,5}

Functional independence occurred in 143/266 versus 139/266, adjusted risk ratio 1.04 (95% CI 0.89–1.20). The adjusted generalised odds ratio for the full mRS distribution was 1.07 (95% CI 0.84–1.35). Neither early neurological change nor quality of life showed a clear benefit. Deaths occurred in 49/266 versus 46/266, 18.4% versus 17.3%; adjusted hazard ratio 1.06 (95% CI 0.71–1.59).⁴

Symptomatic intracranial haemorrhage occurred in 11/267 patients in each group, 4.1%, with an adjusted risk ratio of 1.05 (95% CI 0.45–2.44). Any radiological intracranial haemorrhage occurred in 69/267 versus 66/267, 25.8% versus 24.7%; adjusted risk ratio 1.10 (95% CI 0.82–1.47). Subgroup analyses did not identify convincing evidence of differential benefit. In particular, favourable within-group estimates for women and cardioembolic strokes were not accompanied by clear treatment interactions.⁴

Critique

The trial provides a direct randomised answer to a clinically distinct question. Its strengths include multiple participating centres, allocation after the relevant angiographic result, minimal loss to follow-up, no crossover and blinded functional outcome assessment. These features make it more informative for routine treatment after successful thrombectomy than observational comparisons in which clinicians selected patients for rescue therapy. The primary endpoint also captures a meaningful patient state rather than a purely angiographic surrogate.

The central finding is that this regimen did not establish an improvement in excellent recovery. It does not prove that the true effect is exactly zero. The absolute-effect interval includes modest harm and an important benefit, and the study was powered for a relatively large 13-point improvement. A smaller effect could remain undetected. Nevertheless, the possibility of benefit is a reason for further investigation, not a substitute for evidence when adding treatment to a successful procedure.

The broader outcome pattern matters. The five-point difference at the mRS 0–1 boundary became a 1.5-point difference at the mRS 0–2 boundary, while the full-scale generalised odds ratio was close to one. Mortality did not favour urokinase. This does not invalidate the primary endpoint, but it provides little corroboration for a broad improvement in disability. A claim that the treatment improved recovery because more patients crossed one favourable threshold would therefore be too strong.

Dichotomising the mRS also requires careful language. The trial’s phrase “survival without disability” refers specifically to scores 0–1. Patients at score 1 can still have symptoms, while those at score 2 may remain independent despite being unable to perform all previous activities. It is preferable to define the categories explicitly rather than suggest that all survivors outside the primary-success group are dependent. The clinical value of treatment should also include meaningful improvements among patients who do not reach an excellent outcome.

The population was already selected for a high-quality technical result. Approximately two-thirds had complete angiographic reperfusion. Any additional benefit must therefore come from abnormalities not obvious on the final angiogram or from preventing subsequent deterioration. The trial did not select patients using demonstrated residual tissue hypoperfusion, nor did it require post-drug angiography or perfusion imaging. It consequently cannot determine whether urokinase reached and improved a persistent microvascular target.⁴

This creates two different explanations for a neutral result. The drug may be ineffective in the intended mechanism, or the relevant mechanism may be present in only a minority of otherwise eligible patients. Both remain possible. There are also non-thrombotic causes of poor tissue perfusion and irreversible injury that further fibrinolysis would not reverse. A future imaging-selected trial might be more informative, but selecting on a physiological abnormality would itself require validation and practical delivery within a time-sensitive procedure.

The dose is another unresolved variable. POST-UK tested 100 000 IU for every patient, rather than a weight-adjusted or response-titrated regimen. It cannot identify a dose–response relationship or determine whether a larger dose would improve benefit relative to bleeding. Historical experience with larger amounts used to treat an occluded artery is not a reason to increase the dose after reperfusion. Different clot burdens, exposure times and tissue conditions can alter both efficacy and harm.

The safety findings are less alarming than those of some tenecteplase trials, but should not be labelled proof of safety. Only 22 symptomatic intracranial haemorrhages

occurred. Their confidence interval permits a substantial reduction or more than a doubling of risk. Similarly, the mortality interval includes clinically important effects in either direction. Equal observed event counts do not establish equivalence, particularly for uncommon outcomes. The absence of an excess of all radiological bleeding is useful information, but does not remove this uncertainty.

The bleeding categories also require attention to their definitions. The table headed systemic bleeding uses criteria in which severe bleeding includes intracranial haemorrhage. Those counts should not be interpreted as a separate set of major extracranial bleeds or added to the intracranial events. More generally, adverse-event totals count events rather than necessarily distinct patients. A safety summary that mixes those denominators can substantially overstate the number harmed.^{4,5}

Open-label management remains a potential source of performance bias. The family and clinical team could know whether an adjunctive thrombolytic had been administered, and the control procedure ended without an equivalent infusion. This may influence monitoring, blood-pressure treatment, antithrombotic decisions or subsequent imaging. Central blinded scoring reduces one form of bias, but cannot ensure equivalent co-interventions. The reference to national practice guidance does not provide the same reassurance as detailed evidence of actual care delivered in each group.

The very small amount of missing data strengthens confidence that attrition did not drive the primary result. The different imputation scenarios generated risk ratios around 1.12–1.13, and accounting for centre effects did not reveal a clear benefit. Per-protocol results were also consistent. However, eligibility deviations included prior thrombolysis, longer procedures and less complete reperfusion; removing such patients after allocation should remain a supportive analysis, not the preferred estimate of the treatment policy.⁵

Subgroups do not provide a reliable workaround for the neutral overall result. Women and patients with cardioembolic strokes had favourable within-subgroup estimates, but interaction P values were 0.06 and 0.10, respectively, among multiple comparisons. Even a more extreme interaction would require replication and biological coherence before changing practice. The study does not support routinely treating one sex, one stroke mechanism or one reperfusion category while withholding treatment from another.

Applicability is bounded by the exclusions. Patients who had received intravenous thrombolysis, had substantial preceding disability, large early infarction, difficult prolonged thrombectomy or major bleeding risk were not the intended population. Roughly half had large-artery atherosclerotic stroke, a higher proportion than in many

Western thrombectomy cohorts. These features are more informative than treating recruitment in China as a generic limitation. They specify what clinicians must compare with their own patients and practices.

The companion POST-TNK trial is useful corroborating evidence about the procedural strategy, but not a direct comparison of agents. Its larger excess of radiological bleeding does not prove urokinase is safer, and its numerically lower mortality does not prove tenecteplase is more effective. Separate control groups, differences in stroke mechanisms and chance all remain relevant. A trial randomising the agents within the same population would be needed to answer that comparative question.^{6,7}

Body of Evidence

MELT Japan studied intra-arterial urokinase within six hours in patients with an occluded M1 or M2 artery, before modern thrombectomy became routine. It randomised 114 patients and stopped after intravenous alteplase was approved in Japan. The primary mRS 0–2 endpoint occurred in 49.1% versus 38.6%, odds ratio 1.54 (95% CI 0.73–3.23), without statistical significance. A secondary mRS 0–1 comparison favoured urokinase. This suggested thrombolytic potential, but concerned reopening an artery, not adjunctive treatment after eTICI 2c–3 reperfusion.¹

A single-centre observational cohort subsequently evaluated 100 urokinase-treated patients among 993 anterior-circulation thrombectomy cases. Urokinase was generally selected after failed or incomplete treatment. Among 53 patients with partial or near-complete reperfusion, 18 improved their angiographic grade after administration. Adjusted analyses did not identify increased symptomatic haemorrhage and suggested better functional independence. These findings supported feasibility and further study, but treatment selection and residual confounding prevent them from establishing a causal benefit.²

The INFINITY registry collected 311 recipients of adjunctive alteplase or urokinase across ten European centres. Treatment was predominantly rescue therapy after unsuccessful or incomplete reperfusion. Symptomatic intracranial haemorrhage occurred in 27/308, and angiographic improvement was common among patients with follow-up imaging. Better subsequent reperfusion was associated with better functional outcome. However, the registry lacked a randomised untreated comparator and mixed drugs and indications. It cannot establish that routine urokinase after an already complete result improves outcome.⁸

CHOICE moved the question into a randomised, placebo-controlled setting after successful thrombectomy. Among 113 patients treated as allocated, intra-arterial alteplase produced mRS 0–1 in 36/61 versus 21/52 controls, an adjusted difference of 18.4 percentage points (95% CI 0.3–36.4). The trial allowed eTICI 2b50–3 reperfusion and previous intravenous thrombolysis, and stopped early because of recruitment and placebo-supply difficulties. Its signal is important, but the agent, eligibility and uncertainty differ from POST-UK.³

POST-TNK used the closest parallel design, testing tenecteplase 0.0625 mg/kg after eTICI 2c–3 reperfusion without prior intravenous thrombolysis. Excellent outcomes occurred in 132/269 versus 119/270, adjusted risk ratio 1.15 (95% CI 0.97–1.36). The primary result was neutral and radiological intracranial bleeding increased. Together, the two POST trials do not establish routine low-dose adjunctive fibrinolysis in this narrow population, although neither excludes a modest benefit.⁶

Later 2025 trials broadened the evidence. ANGEL-TNK used tenecteplase 0.125 mg/kg after eTICI 2b–3 reperfusion in the 4.5–24-hour window. Its primary mRS 0–1 result favoured treatment, 51/126 versus 34/129, reported adjusted relative risk 1.44 (95% CI 1.06–1.95), but secondary efficacy analyses did not establish benefit. PEARL tested alteplase 0.225 mg/kg, allowed prior intravenous thrombolysis and eTICI 2b50–3 reperfusion, and found mRS 0–1 in 73/163 versus 48/159, adjusted risk ratio 1.45 (95% CI 1.08–1.96). Mortality and radiological bleeding were numerically higher with alteplase, with uncertainty around both.^{9,10}

These positive primary results prevent a blanket conclusion that all adjunctive intra-arterial thrombolysis is ineffective. They also do not validate urokinase 100 000 IU after near-complete reperfusion. Pooling all trials as a single drug-class effect requires assumptions about agent, dose, previous thrombolysis, residual reperfusion and patient selection. The joint POST editorial's call for further comparative evidence remains relevant, while its early-2025 synthesis should be read alongside the later trials.⁷

POST-UK's contribution is therefore specific and useful: an accessible low-dose urokinase strategy did not demonstrate better excellent recovery, broader functional improvement or survival after near-complete or complete thrombectomy reperfusion. Routine use is unsupported. Rescue treatment for persistent occlusion, alternative dosing and selection by residual tissue hypoperfusion remain distinct questions rather than benefits that can be inferred from this trial.

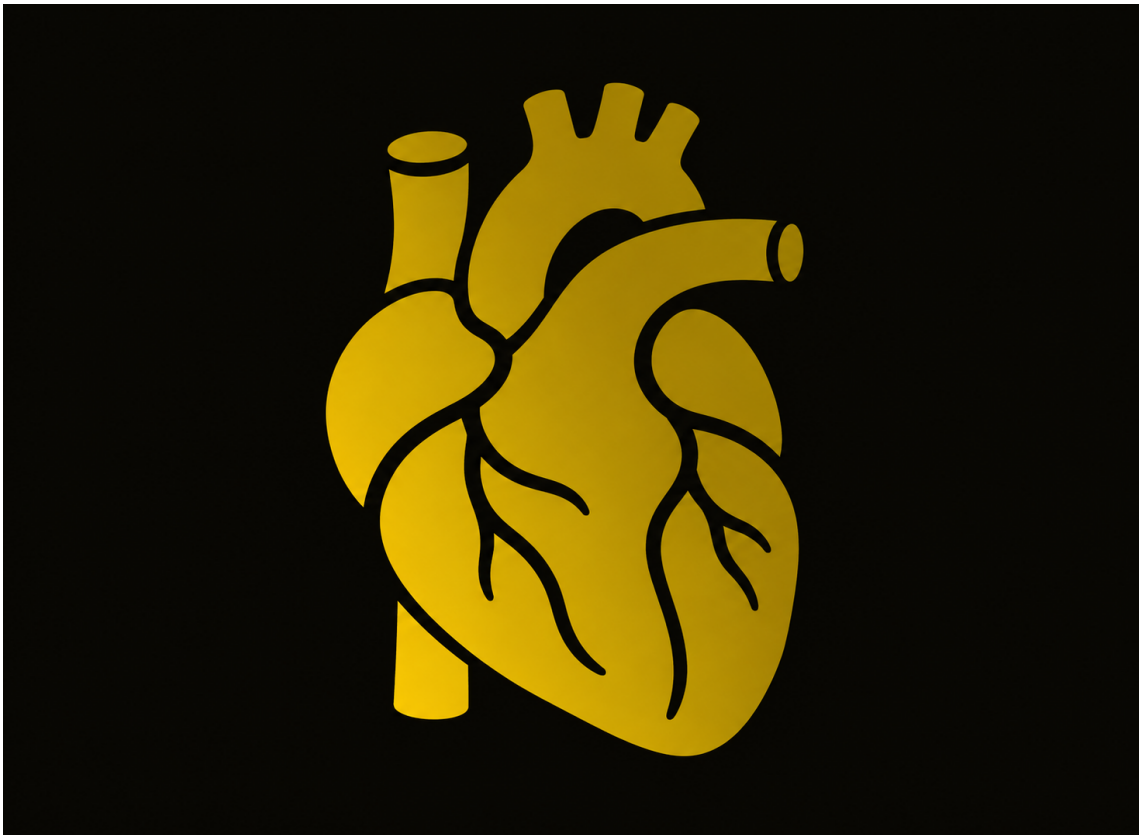
Should 100 000 IU of intra-arterial urokinase routinely follow successful thrombectomy?

No: POST-UK did not establish improved functional recovery after eTICI 2c–3 reperfusion. Similar observed bleeding rates do not prove safety equivalence or superiority over other thrombolytics.

References

1. Ogawa A, Mori E, Minematsu K, Taki W, Takahashi A, Nemoto S, et al. Randomized trial of intraarterial infusion of urokinase within 6 hours of middle cerebral artery stroke: the middle cerebral artery embolism local fibrinolytic intervention trial (MELT) Japan. *Stroke*. 2007;38(10):2633-2639.
2. Kaesmacher J, Bellwald S, Dobrocky T, Meinel TR, Piechowiak EI, Goeldlin M, et al. Safety and Efficacy of Intra-arterial Urokinase After Failed, Unsuccessful, or Incomplete Mechanical Thrombectomy in Anterior Circulation Large-Vessel Occlusion Stroke. *JAMA Neurol*. 2020;77(3):318-326.
3. Renú A, Millán M, San Román L, Blasco J, Martí-Fàbregas J, Terceño M, et al. Effect of Intra-arterial Alteplase vs Placebo Following Successful Thrombectomy on Functional Outcomes in Patients With Large Vessel Occlusion Acute Ischemic Stroke: The CHOICE Randomized Clinical Trial. *JAMA*. 2022;327(9):826-835.
4. Liu C, Guo C, Li F, Yu N, Huang J, Peng Z, et al. Intra-Arterial Urokinase After Endovascular Reperfusion for Acute Ischemic Stroke: The POST-UK Randomized Clinical Trial. *JAMA*. 2025;333(7):589-598.
5. Liu C, Guo C, Li F, Yu N, Huang J, Peng Z, et al. Supplementary material to: Intra-Arterial Urokinase After Endovascular Reperfusion for Acute Ischemic Stroke: The POST-UK Randomized Clinical Trial. *JAMA*. 2025;333(7):589–598.
6. Huang J, Yang J, Liu C, Li L, Yang D, Guo C, et al. Intra-Arterial Tenecteplase Following Endovascular Reperfusion for Large Vessel Occlusion Acute Ischemic Stroke: The POST-TNK Randomized Clinical Trial. *JAMA*. 2025;333(7):579-588.
7. Dippel DWJ, Khan CF, Schoon BA. Intra-Arterial Thrombolytics During Thrombectomy for Ischemic Stroke-End of the Story or a New Beginning?. *JAMA*. 2025;333(7):571-573.

8. Kaesmacher J, Abdullayev N, Maamari B, Dobrocky T, Vynckier J, Piechowiak EI, et al. Safety and Angiographic Efficacy of Intra-Arterial Fibrinolytics as Adjunct to Mechanical Thrombectomy: Results from the INFINITY Registry. *J Stroke*. 2021;23(1):91-102.
9. Miao Z, Luo G, Song L, Sun D, Chen W, Yao X, et al. Intra-arterial Tenecteplase for Acute Stroke After Successful Endovascular Therapy: The ANGEL-TNK Randomized Clinical Trial. *JAMA*. 2025;334(7):582-591.
10. Writing Committee for the PEARL Investigators. Intra-Arterial Alteplase After Successful Endovascular Reperfusion in Acute Stroke: The PEARL Randomized Clinical Trial. *JAMA*. 2025;334(19):1728–1739.



Circulatory Trials

Altshock-2

Morici N, Sacco A, Frea S, Rota M, Villanova L, Gravinese C, et al. Early Intra-Aortic Balloon Support for Heart Failure-Related Cardiogenic Shock: A Randomized Clinical Trial. J Am Coll Cardiol. 2025;85(16):1587-1597.

Introduction

Cardiogenic shock complicating decompensated chronic heart failure presents a different therapeutic problem from shock following an acute myocardial infarction. The failing ventricle may be markedly dilated, congested and sensitive to afterload, with low cardiac output developing against a background of longstanding adaptation. Recovery with medical treatment, transplantation and durable ventricular assistance are all possible destinations. Temporary circulatory support must therefore be judged by whether it helps patients reach a worthwhile destination, as well as by the immediate haemodynamic changes it produces.

An intra-aortic balloon pump inflates during diastole and deflates before systole, augmenting aortic diastolic pressure and reducing left ventricular afterload. Its circulatory assistance is modest, but placement is familiar and generally requires a smaller arterial access than more powerful devices. In heart failure with substantial afterload sensitivity, reducing the resistance against which the ventricle ejects could be useful even without providing several litres of additional flow. This physiological argument does not itself establish clinical benefit.

The absence of a mortality benefit in IABP-SHOCK II concerned infarct-related shock and could not settle the question in decompensated heart failure. A small randomised comparison with inotropes suggested a physiological benefit in a more specifically selected low-output population, but provided little precision for patient-centred outcomes. Altshock-2 therefore examined whether placing a balloon pump routinely and early improved survival or successful bridging to heart replacement therapy in patients with heart failure-related cardiogenic shock.^{1,2,3}

Synopsis

Altshock-2 was an open-label, multicentre randomised trial conducted at seven Italian centres between June 2020 and October 2024. An unrestricted grant from Getinge supported the trial; the funder did not participate in design, analysis, interpretation or writing. Clinical events were assessed by an independent committee masked to treatment allocation. A separate monitoring board reviewed safety and the planned interim analysis.^{3,4}

Patients were aged 18–75 years, had heart failure-related cardiogenic shock and were considered potentially suitable for transplantation or a durable left ventricular assist device. Acute myocardial infarction within the previous month, myocarditis, pulmonary embolism, postoperative shock, septic shock, severe peripheral vascular disease and irreversible extracardiac disease precluding heart replacement were important exclusions. The trial consequently addressed a selected advanced heart failure population rather than all causes of cardiogenic shock.

Eligibility initially required hypotension and hypoperfusion. An amendment approved in December 2022 broadened recruitment to Society for Cardiovascular Angiography and Interventions stages B–D. Stage B denotes an earlier, less overtly hypoperfused state than established shock. Twelve patients without hypoperfusion had already been enrolled as deviations before that amendment. Although deferred consent could be used where locally approved, all participants ultimately enrolled gave consent before randomisation. Fifty-two screened patients were excluded because consent was unavailable.

Central web-based allocation used variable permuted blocks to assign patients equally to immediate balloon-pump insertion plus medical treatment or medical treatment initially without a balloon pump. Clinicians and patients knew the allocation. Either of the available balloon-pump systems could be used, with catheter size determined by patient height and placement at the bedside or in the catheter laboratory. Pulmonary artery catheterisation was encouraged but not required. Prophylactic heparin was recommended in the intervention group unless another indication required a different antithrombotic regimen.

Both groups received locally determined medical treatment. Drug titration, balloon-pump optimisation and weaning were not standardised in detail. Inotropic support was limited to a score of 20 before mandatory escalation. Further respiratory or circulatory support was considered if patients failed to meet at least five of eight goals by eight hours: heart rate below 100/min, mean arterial pressure above 65 mmHg, mixed venous oxygen saturation above 60%, arterial oxygen tension above 60 mmHg, lactate

below 3 mmol/L, low filling pressure, an inotropic score below 20 and urine output above 0.5 mL/kg/h. Crossover was discouraged but rescue balloon-pump insertion remained possible.

The primary outcome was survival at 60 days with either native heart recovery or successful bridging to transplantation or durable ventricular assistance. The report presented the proportion reaching this favourable outcome and a time-to-event hazard ratio for the adverse outcome. Secondary measures included maximum inotropic requirement, duration of inotropic therapy and maximum Sequential Organ Failure Assessment score. Bleeding, vascular complications and embolic events were assessed as harms.

Planning assumed a favourable primary outcome in 75% with balloon support and 55% with standard care, an absolute difference of 20 percentage points. Two hundred participants would provide 80% power with a two-sided significance level of 5%. An interim review was scheduled after approximately half the target had completed follow-up. The principal analysis followed randomised allocation and used an unadjusted Cox model. Secondary comparisons and exploratory subgroups did not provide an independent confirmatory framework for smaller treatment effects.

Of 787 screened patients, 101 were randomised: 53 to early balloon support and 48 to standard care. All were retained in the principal analysis. Median age was 60 years, 80% were men and median left ventricular ejection fraction was approximately 20%. Median lactate was only 1.85 mmol/L. Twenty-eight patients were classified as stage B; most of the remainder were stage C and a smaller group stage D. Right ventricular dysfunction, defined by reduced tricuspid annular excursion, was present in about 70%. Approximately 60% had non-ischaemic cardiomyopathy, and many already had an implanted defibrillator or resynchronisation device.

Balloon insertion succeeded in every allocated patient. Six controls, 13% of that group, subsequently received a balloon pump. Escalation to a more powerful temporary support device occurred in four intervention patients and two controls. Pulmonary artery catheters were used in 24 of 53 intervention patients and 20 of 48 controls. Inotropes were given to 64% and 69%, respectively, while vasodilators were used in 70% and 65%. The comparison was therefore an early device strategy added to active haemodynamic treatment, with rescue available in the control arm.

At the interim review, the monitoring board judged that completing recruitment was unlikely to yield the intended benefit and recommended stopping for futility. The final sample of 101 was approximately half the planned total. The main publication describes

the timing and judgement behind this decision but does not supply a numerical conditional-power calculation supporting it.

The favourable primary outcome occurred in 43 of 53 patients assigned to early support, 81%, and 36 of 48 controls, 75%. The reported hazard ratio was 0.72, with a 95% confidence interval of 0.31–1.68 and $P=0.45$. The corresponding adverse-outcome counts were ten and twelve. The hazard ratio should not be interpreted as a ratio of the favourable percentages. A total of 37 patients underwent heart replacement therapy: 21 in the early-support arm and 16 controls. Across both groups, 18 underwent transplantation and 19 received a durable ventricular assist device; seven subsequently died after these procedures.

There was no clear separation in maximum inotropic score, inotrope duration, organ failure or length of hospital stay. Severe bleeding occurred in nine of 53 intervention patients, 17%, and four of 48 controls, 8.3%. Vascular complications occurred in four intervention patients and no controls; limb ischaemia occurred in two and stroke in three intervention patients. These small counts did not produce statistically conclusive between-group differences. An analysis labelled per protocol reassigned the six crossover patients to the balloon-pump group and gave a hazard ratio of 0.99, with a 95% confidence interval of 0.42–2.31.³

Critique

Altshock-2 failed to demonstrate that routine early balloon support improves its selected clinical outcome. It did not demonstrate that the two strategies are equivalent. Only 22 patients experienced the adverse primary outcome, and the confidence interval encompasses a substantial reduction or increase in its hazard. The appropriate inference is that routine insertion lacks demonstrated benefit in this population, while the magnitude of any smaller effect remains uncertain.

The original assumptions made a large benefit relatively easy to detect if it existed: a rise in favourable outcomes from 55% to 75%. In practice, controls achieved 75%. This reduced the scope for the anticipated absolute gain and made the enrolled population different from the one implicit in the power calculation. Stopping at roughly half the planned sample further limited precision. A futility decision can be reasonable for the planned hypothesis without establishing that every clinically worthwhile effect is absent. The declining recruitment rate also formed part of the circumstances in which the board considered continuation.

Disease severity is central to applicability. A median lactate below 2 mmol/L, modest inotropic requirements and inclusion of stage B patients describe a population quite different from rapidly deteriorating shock with escalating vasopressors, severe acidosis or cardiac arrest. All enrolled patients could provide consent, and all were potential candidates for definitive heart replacement. These features help explain the favourable overall outcome without proving that low severity caused the neutral comparison. They also prevent extrapolation to the older, frailer or profoundly unstable patients commonly encountered in unselected critical care.

The eligibility amendment introduces another layer of heterogeneity. Earlier recruitment targeted overt hypoperfusion, whereas later recruitment explicitly allowed an earlier stage. The inclusion of twelve such patients before the amendment is a documented departure from the initial criteria. Randomisation still protects the comparison across the enrolled cohort, but it does not convert that cohort into a precise test of one haemodynamic phenotype. The small stage subgroups cannot reliably identify a population that benefits.

A particularly relevant unanswered question is whether benefit requires demonstrable afterload mismatch. Neither high systemic vascular resistance nor a specific combination of low cardiac index and high filling pressure was mandatory. Fewer than half the participants underwent pulmonary artery catheterisation. Diastolic augmentation and other measures of device performance were not systematically collected. Correspondence highlighted these omissions and proposed more physiological selection. That is a defensible research hypothesis, but it is not evidence that selecting such patients would have reversed the trial result.⁵

The treatment contrast also deserves a precise description. This was not balloon support against inotropes alone, nor balloon support against a prohibition on later mechanical assistance. Both groups received varied combinations of drugs, including frequent vasodilator use, and controls could receive rescue support. A device that principally reduces afterload might add less when medical afterload reduction is already effective. However, treatment choices were not independently randomised, and their contribution to the observed result cannot be isolated.

Crossover weakens the contrast between actually receiving and never receiving a balloon pump, but it does not invalidate the principal pragmatic question. The intention-to-treat estimate concerns adopting routine early insertion rather than reserving insertion for selected subsequent deterioration. Reassigning the six controls who later received support does not recover an unbiased device effect: deterioration itself can influence both crossover and prognosis. The analysis described as per protocol

is therefore more accurately understood as a treatment-reclassified sensitivity analysis, not independent confirmation of equivalence.

The primary endpoint combines different clinical pathways. Surviving with the native heart and surviving after transplantation or durable support are all meaningful, but they involve very different burdens and depend partly on candidacy, organ availability and clinician decisions. Blinded event adjudication strengthens classification, yet cannot blind decisions to list, implant or escalate. A 60-day horizon may also miss later recovery or durable-device implantation. These limitations matter most when applying the result to systems with different access to advanced heart failure therapies.

Safety requires the same restraint as efficacy. Bleeding was numerically more frequent with early support, and vascular complications and strokes occurred only in that allocated group. The sample is too small to estimate these risks reliably or establish causation for each event. Describing complications as statistically similar must not become a claim that routine balloon insertion is harmless. In the absence of established benefit, even an uncertain procedural burden weighs against automatic use.

The trial's main contribution is the randomised examination of a practice previously supported largely by physiology and clinical selection. Its findings favour continued assessment and medical treatment before routine insertion in comparable patients. They leave room for rescue decisions in individual deterioration, but do not identify the ideal escalation threshold, prove a preferred alternative device or establish which physiological subgroup should receive a balloon pump.

Body of Evidence

IABP-SHOCK II randomised 600 patients with infarct-related cardiogenic shock to balloon support or no planned balloon support alongside early revascularisation and medical treatment. Thirty-day mortality was 39.7% versus 41.3%, with a relative risk of 0.96 and a 95% confidence interval of 0.79–1.17. This removed the expectation that routine counterpulsation necessarily improves survival in acute infarct-related shock. Its relevance to Altshock-2 is conceptual rather than directly interchangeable: acute coronary occlusion, reperfusion and myocardial injury differ from decompensation of a chronically failing, afterload-sensitive ventricle.¹

Den Uil and colleagues provided a more closely related but much smaller randomised comparison. Thirty-two patients with decompensated heart failure and low output received primary balloon support or inotropes. The trial selected patients using invasive haemodynamic evidence and focused on the early change in mixed venous oxygen

saturation. Balloon support improved this physiological measure more than inotropes, while the small sample could not establish a mortality benefit. It supported biological plausibility in a selected low-output state; it did not establish that routine balloon insertion improves survival across the broader population recruited in Altshock-2.²

More powerful devices cannot be treated as interchangeable substitutes on the basis of these results. ECLS-SHOCK evaluated early extracorporeal life support in 417 analysed patients with infarct-related shock undergoing planned revascularisation. Thirty-day mortality was 47.8% with extracorporeal support and 49.0% with usual care, without a significant benefit, while bleeding and vascular complications increased. Its result illustrates the importance of balancing circulatory assistance against the harms required to deliver it, but does not directly compare extracorporeal support with a balloon pump in chronic heart failure.⁶

Conversely, DanGer-Shock found a survival benefit from a microaxial flow pump in a selected infarct-related shock population. Among 355 analysed patients, 180-day mortality was 45.8% with the pump and 58.5% with standard care, with a hazard ratio of 0.74 and a 95% confidence interval of 0.55–0.99. Major device-related safety events and renal replacement therapy were more frequent. The contrasting result reinforces the importance of population, device and treatment pathway; it does not justify assuming that a stronger pump would have improved the outcome in Altshock-2.⁷

Taken together, these trials argue against treating temporary mechanical support as a single therapy with a uniform effect. Altshock-2 adds direct evidence in an advanced heart failure population, but its breadth of haemodynamic selection and limited precision leave the selective-use question open. Future comparisons need to connect demonstrable physiological need, the timing of escalation and the intended destination of support to outcomes that matter to patients.

Should we routinely insert an early intra-aortic balloon pump in heart failure-related cardiogenic shock?

Not routinely in patients resembling those enrolled in Altshock-2. The trial found no clear benefit, while selective rescue use and more precisely defined haemodynamic subgroups remain unresolved.

References

1. Thiele H, Zeymer U, Neumann FJ, Ferenc M, Olbrich HG, Hausleiter J, et al. Intraaortic balloon support for myocardial infarction with cardiogenic shock. *N Engl J Med*. 2012;367(14):1287-1296.
2. den Uil CA, Van Mieghem NM, B Bastos M, Jewbali LS, Lenzen MJ, Engstrom AE, et al. Primary intra-aortic balloon support versus inotropes for decompensated heart failure and low output: a randomised trial. *EuroIntervention*. 2019;15(7):586-593.
3. Morici N, Sacco A, Frea S, Rota M, Villanova L, Gravinese C, et al. Early Intra-Aortic Balloon Support for Heart Failure-Related Cardiogenic Shock: A Randomized Clinical Trial. *J Am Coll Cardiol*. 2025;85(16):1587-1597.
4. Morici N, Sacco A, Frea S, Rota M, Villanova L, Gravinese C, et al. Supplementary appendix to Early Intra-Aortic Balloon Support for Heart Failure-Related Cardiogenic Shock: A Randomized Clinical Trial. *J Am Coll Cardiol*. 2025.
5. Baldetti L, Ajello S, Gramegna M, den Uil CA, Scandroglio AM. Intra-Aortic Balloon Pump in Heart Failure-Related Cardiogenic Shock: Dead or Alive? *J Am Coll Cardiol*. 2025;86(4):e45-e46.
6. Thiele H, Zeymer U, Akin I, Behnes M, Rassaf T, Mahabadi AA, et al. Extracorporeal Life Support in Infarct-Related Cardiogenic Shock. *N Engl J Med*. 2023;389(14):1286-1297.
7. Møller JE, Engstrøm T, Jensen LO, Eiskjær H, Mangner N, Polzin A, et al. Microaxial Flow Pump or Standard Care in Infarct-Related Cardiogenic Shock. *N Engl J Med*. 2024;390(15):1382-1393.

LEVOECMO

Combes A, Saura O, Nessler N, Lebbah S, Rozec B, Levy B, et al. Levosimendan to Facilitate Weaning From ECMO in Patients With Severe Cardiogenic Shock: The LEVOECMO Randomized Clinical Trial. JAMA. 2026;335(1):60-69. Published online December 1, 2025.

Introduction

Venoarterial extracorporeal membrane oxygenation can sustain circulation while a severely impaired heart recovers, but continued support brings substantial burdens. Bleeding, thrombosis, limb ischaemia, infection and left ventricular distension can complicate an already precarious course. A treatment that safely accelerates myocardial recovery could shorten exposure to the circuit and help patients survive with their native heart. Simply removing ECMO earlier, however, is not enough if subsequent death or renewed mechanical support follows.

Levosimendan offers an attractive physiological hypothesis. It sensitises contractile proteins to calcium and causes vasodilatation, potentially increasing forward flow while reducing ventricular loading. An active metabolite prolongs its effects beyond a 24-hour infusion. Observational studies in patients receiving venoarterial ECMO suggested improved weaning, although treatment selection, timing and inconsistent definitions made causal interpretation difficult. Patients judged likely to recover may be preferentially offered a drug intended to facilitate recovery, while others receive it only after failed attempts at liberation. These opposite selection mechanisms cannot be fully removed by simple comparisons.^{1,2}

Large randomised trials in cardiac surgery and sepsis had already shown that favourable haemodynamic reasoning does not guarantee improved clinical outcomes with levosimendan. Those results were indirect for ECMO, where the circulation is mechanically supported and the practical objective is sustained separation from the device. LEVOECMO supplied a direct, masked comparison of early levosimendan with placebo in severe but potentially reversible cardiogenic shock.^{3,4,5,6}

Synopsis

LEVOECMO was an investigator-initiated, double-blind, placebo-controlled trial conducted at 11 experienced French intensive care units. Recruitment ran from August 2021 to September 2024, with final follow-up in November 2024. The French Ministry of Health funded the study and Assistance Publique–Hôpitaux de Paris sponsored and coordinated it, including monitoring and statistical analysis. Orion supplied study preparations initially and later contributed to the cost of levosimendan. The first report appeared online on 1 December 2025.^{6,7}

Adults were eligible if they had acute cardiogenic shock refractory to conventional therapy and had started venoarterial ECMO within the preceding 48 hours. The treating team had already made the decision to initiate ECMO. The trial therefore did not evaluate whether ECMO itself should be used.

Important exclusions included prolonged resuscitation exceeding 30 minutes, irreversible neurological injury, end-stage cardiomyopathy without an expectation of recovery, mechanical complications of infarction, more than moderate aortic regurgitation, previous heart transplantation and a moribund state. Patients with a Simplified Acute Physiology Score II above 90, severe hepatic disease or chronic dialysis dependence were also excluded. The supplementary criteria excluded ECMO for pulmonary embolism or cardiotoxic poisoning. Consent could come from a representative, with emergency inclusion permitted when one was unavailable and continued participation discussed subsequently.

A concealed central internet service assigned participants equally to levosimendan or placebo. After an initial balanced random allocation, a stochastic minimisation procedure balanced centre and shock aetiology: postcardiotomy, acute myocardial infarction, myocarditis or another cause. Patients, clinicians and outcome assessors remained masked. Placebo contained riboflavin to reproduce the yellow appearance of levosimendan. During the later recruitment phase, a designated nurse preparing treatment knew which preparation was being used, while the clinical and assessment teams remained masked.

The intervention was a 24-hour intravenous infusion without a loading bolus. The starting rate was 0.15 micrograms/kg/min, rising after two hours to 0.20 micrograms/kg/min if tolerated. The infusion could stop for anaphylaxis, intractable arrhythmia or severe refractory hypotension. The placebo arm followed the same infusion schedule. Both groups received similar approaches to sedation, anticoagulation, haemodynamic management and ECMO care, with a shared weaning procedure. Background inotropes and left ventricular unloading were permitted.

The primary outcome was time to successful ECMO removal within 30 days of randomisation. Success required the patient to remain alive without a new ECMO run, another mechanical circulatory support device or transplantation for 30 days after removal. Thus, a patient decannulated on day 25 required observation to day 55 before that removal could be classified as successful, although the event time remained day 25.

Death while receiving ECMO and failure after removal were competing outcomes. The latter included death, renewed ECMO, another support device or transplantation within the subsequent 30 days. A patient alive but still receiving ECMO on day 30 was censored. This framework distinguished sustained native-heart recovery from device removal followed by another adverse pathway.

Secondary outcomes included mortality at 30 and 60 days, organ failure-free days, duration of ECMO, ventilation and vasoactive treatment, major cardiovascular events, and intensive care and hospital stay. For relevant days-alive-and-free outcomes, death was assigned zero days. Ventricular and supraventricular arrhythmias, severe hypotension and other suspected drug effects were assessed. A planned time-to-haemodynamic-improvement outcome was not analysed because most patients were already haemodynamically stable on support; blood pressure trajectories were presented instead.

The sample-size calculation assumed successful weaning in 50% of controls and 70% with levosimendan, corresponding to a subdistribution hazard ratio of 1.75. A total of 206 participants would provide 80% power at a two-sided 5% significance level. The primary intention-to-treat analysis used competing-risk methods and a Fine–Gray model. A cause-specific hazard analysis was a sensitivity analysis. Secondary endpoints were not supported by a multiplicity-adjusted hierarchy, and the principal results table withheld their P values for that reason.

The study assessed 209 patients for eligibility and randomised 205: 101 to levosimendan and 104 to placebo. All allocated levosimendan patients received treatment; two placebo patients did not. Two intervention patients were lost to follow-up, one before collection of the primary endpoint and one afterwards. All 205 remained in the analysis, with censoring where necessary. Median time from ECMO initiation to randomisation was 25 hours.

Median age was 58 years and 73% were men. Postcardiotomy shock accounted for 79 patients, acute myocardial infarction for 56 and myocarditis for 28. Median Sequential Organ Failure Assessment score was 12 and measured left ventricular ejection fraction approximately 15%. By randomisation, median lactate was approximately 2 mmol/L and

mean arterial pressure approximately 75 mmHg. These values describe patients after initial ECMO stabilisation, not their physiology when rescue support began. Around four-fifths received dobutamine and approximately 72% noradrenaline. Balloon-pump unloading was used in about 37%, with a microaxial pump in about 5%.

The intended higher infusion rate was reached in 93% of levosimendan patients and 96% of placebo patients. Definitive interruption before 24 hours occurred in eleven versus three patients. Successful weaning occurred in 69 of 101 versus 71 of 104, 68.3% in both groups. The absolute difference was 0.0 percentage points, with a 95% confidence interval of -12.8 to 12.7. The subdistribution hazard ratio was 1.02, with a 95% confidence interval of 0.74–1.39 and $P=0.92$. The cause-specific sensitivity estimate was also inconclusive, at 1.15 with a 95% confidence interval of 0.82–1.62.^{6,7}

There were fifteen versus twelve deaths before ECMO removal and fifteen versus twenty-one classified weaning failures. One levosimendan patient remained on ECMO at day 30 and was censored, alongside the patient lost before an outcome. At 60 days, 28 of 101 versus 26 of 104 had died, 27.7% versus 25.0%; the risk difference was 2.7 percentage points, with a 95% confidence interval of -9.0 to 15.3. Median ECMO duration was five versus six days, with an estimated median difference of -1 day and a 95% confidence interval of -2 to 1.

There was no consistent benefit across organ support or intensive care outcomes. Restricted mean hospital stay through day 60 was shorter with levosimendan, 28 versus 35 days, but intensive care stay was similar. Ventricular tachycardia or fibrillation occurred in eighteen versus nine patients, 17.8% versus 8.7%. The reported absolute risk difference was 9.2 percentage points, with a 95% confidence interval of 0.4–18.1; the relative risk was 2.06, with a 95% confidence interval of 0.97–4.37. Four versus one required electrical treatment for ventricular arrhythmia. No convincing treatment interaction was demonstrated across shock aetiologies.

Critique

LEVOECMO provides no evidence that routine early levosimendan accelerates sustained liberation from venoarterial ECMO. The point estimate is almost exactly neutral, actual success proportions are identical to one decimal place, and related outcomes do not supply a coherent alternative benefit. This is a persuasive result against adopting the tested routine strategy, even though the trial cannot exclude smaller effects or all benefits in narrowly selected patients.

The primary endpoint is a major strength. An early decannulation followed by death, renewed support or transplantation is not counted as uncomplicated recovery. The required 30-day period after removal protects against mistaking a transient procedural success for durable liberation. It also makes the endpoint more demanding than several earlier observational definitions, so apparently discordant weaning percentages across studies may reflect different outcomes as well as different treatment effects.

The competing-risk analysis is appropriate to the clinical pathways being observed. Death can prevent successful weaning, and removal followed by failure is another outcome, not simply missing follow-up. The subdistribution hazard ratio summarises the treatment effect on cumulative successful weaning in the presence of those competing events; it is not a ratio of median ECMO durations or a direct measure of myocardial contractility. The cause-specific analysis offers a different statistical view, but does not by itself establish a purely biological effect with competing outcomes magically removed.

There is still a limit to what the endpoint values. Transplantation or durable ventricular assistance may be a successful clinical rescue, yet counts as failure of native-heart liberation. That definition fits a study selecting patients with potential for myocardial recovery, but it should not imply that all alternative destinations are equally undesirable. Reporting mortality and the separate pathways therefore remains essential. Here, neither survival nor the broader cardiovascular composite established an offsetting benefit.

The planned treatment effect was large. Control weaning success was 68.3%, much higher than the assumed 50%, leaving less scope for a 20-percentage-point absolute improvement. The confidence interval around the risk difference still allows benefits or harms of approximately thirteen percentage points. It would be wrong to declare equivalence. Equally, an appeal to limited power does not turn the near-null estimate into a positive clinical rationale: a treatment with possible adverse effects requires demonstrated benefit before routine addition.

Treatment delivery was substantial. Most patients received the intended rate of 0.20 micrograms/kg/min, rather than an unusually low or brief exposure. More intervention infusions were stopped early, which is itself part of the tolerability of the strategy. The result cannot readily be dismissed as a failure to administer the drug. However, this was a single early infusion, usually around a day after ECMO initiation; it did not test targeted use after a failed weaning attempt or repeated dosing selected by a physiological response.

Patient selection also matters. Severe irreversible neurological injury, prolonged resuscitation, advanced organ disease and end-stage cardiomyopathy were excluded. The high rate of eventual weaning and relatively low mortality therefore should not be assumed to apply to every ECMO population. Conversely, the normalised lactate and supported blood pressure at enrolment do not mean the patients had never been critically ill. They reflect a trial of recovery after mechanical stabilisation, rather than a trial of initial resuscitation.

Background treatment was considerable: most patients received conventional inotropes and many received mechanical left ventricular unloading. Levosimendan was added to this treatment, not substituted for all catecholamines or unloading devices. A potential physiological effect could be less visible in a pathway already supplying these functions, but the pragmatic question remains whether adding the drug improves outcomes. The trial gives no evidence that it does.

The myocarditis subgroup is intriguing but insufficient for practice. Eleven of twelve levosimendan patients and ten of sixteen controls achieved successful weaning. These small numbers produce wide uncertainty, and the interaction across aetiologies was not significant. A favourable subgroup estimate within an overall neutral trial is not proof of a treatment-responsive phenotype. It identifies a question that would need a suitably designed comparison.

The arrhythmia signal is clinically relevant. Eighteen versus nine ventricular events is compatible with a meaningful increase, but precision is limited and secondary analyses were numerous. The absolute-difference interval excludes zero whereas the relative-risk interval narrowly includes one, reflecting different estimation methods rather than two independent experiments. The appropriate description is concern about ventricular arrhythmias, not a claim of a precisely doubled proven risk. Greater treatment cessation reinforces the need to consider tolerability alongside efficacy.

Shorter hospital stay is not a sufficient counterargument. Death can shorten time in hospital, and the groups differed numerically in mortality and transplantation. The endpoint was restricted mean observed hospital duration through 60 days, not necessarily faster discharge alive with comparable health. Without improvement in successful weaning, intensive care stay or survival, this isolated secondary difference should not drive adoption.

Finally, several supplementary rows label ventilation and kidney-support durations and free days in ways that appear reversed. Those values should not be silently corrected or used to claim an organ-specific effect. The primary endpoint, mortality and main-table

ECMO data are sufficiently clear to support the central interpretation while the detailed secondary labels await clarification. The later change in drug preparation also deserves acknowledgement, but awareness by a designated preparation nurse does not establish that clinical masking failed.

Body of Evidence

The observational ECMO literature initially encouraged use. Vally and colleagues' retrospective study compared propensity-matched patients and associated levosimendan with less weaning failure, but its adjusted mortality estimate remained imprecise. A subsequent meta-analysis of fifteen publications involving 1,772 patients found higher weaning success, with a pooled odds ratio of 2.78, but substantial heterogeneity. Combining selected observational cohorts cannot eliminate unmeasured differences in recovery potential, treatment timing or endpoint definition. These reports supplied a rationale for randomisation, not a secure estimate of benefit.^{1,2}

A 2025 bicentre propensity-matched analysis by Paulo and colleagues used a more rigorous definition requiring survival without transplantation or durable support after removal. Among 47 matched pairs, successful weaning occurred in 45% with levosimendan and 34% without it, with a subdistribution hazard ratio of 1.45 and a 95% confidence interval of 0.77–2.70. The study excluded postcardiotomy shock and remained observational. Its uncertainty and lack of a clear clinical advantage foreshadowed the need for LEVOECMO rather than resolving the issue.⁸

CHEETAH tested levosimendan after cardiac surgery when haemodynamic support was required. It stopped for futility after 506 patients, with 30-day mortality of 12.9% versus 12.8% on placebo. LEVO-CTS examined prophylactic treatment in patients with impaired left ventricular function undergoing surgery: among 849 treated patients, the four-component primary outcome occurred in 24.5% of each group. These are different settings and dosing pathways, but both caution against inferring clinical benefit from an inodilator's physiological appeal.^{3,4}

LeoPARDS supplies further indirect evidence. In 516 patients with sepsis, levosimendan did not improve mean organ failure score and was associated with more supraventricular tachyarrhythmia and a lower likelihood of successful ventilator liberation. Septic shock is not equivalent to predominantly cardiac failure on ECMO, so its result cannot substitute for the direct trial. It does reinforce that calcium sensitisation and vasodilatation can carry clinically important costs.⁵

LEVOECMO now places a randomised, masked comparison at the centre of this specific decision. Earlier associations with easier weaning have not translated into demonstrated benefit from a routine early infusion. Selective rescue use after failed liberation remains untested, but the burden of proof for that approach cannot be met by reinterpreting the neutral overall trial or one small subgroup.

Should we routinely give levosimendan early during venoarterial ECMO to facilitate weaning?

No. LEVOECMO found no improvement in sustained successful weaning and raised concern about ventricular arrhythmias.

References

1. Vally S, Ferdynus C, Persichini R, Bouchet B, Braunberger E, Lo Pinto H, et al. Impact of levosimendan on weaning from peripheral venoarterial extracorporeal membrane oxygenation in intensive care unit. *Ann Intensive Care*. 2019;9(1):24.
2. Liu Y, Zhang L, Yao Y, Li Y, Qin W, Li Y, et al. Effects of levosimendan on the outcome of veno-arterial extracorporeal membrane oxygenation: a systematic review and meta-analysis. *Clin Res Cardiol*. 2024;113(4):509-521.
3. Landoni G, Lomivorotov VV, Alvaro G, Lobreglio R, Pisano A, Guarracino F, et al. Levosimendan for Hemodynamic Support after Cardiac Surgery. *N Engl J Med*. 2017;376(21):2021-2031.
4. Mehta RH, Leimberger JD, van Diepen S, Meza J, Wang A, Jankowich R, et al. Levosimendan in Patients with Left Ventricular Dysfunction Undergoing Cardiac Surgery. *N Engl J Med*. 2017;376(21):2032-2042.
5. Gordon AC, Perkins GD, Singer M, McAuley DF, Orme RM, Santhakumaran S, et al. Levosimendan for the Prevention of Acute Organ Dysfunction in Sepsis. *N Engl J Med*. 2016;375(17):1638-1648.
6. Combes A, Saura O, Nessler N, Lebbah S, Rozec B, Levy B, et al. Levosimendan to Facilitate Weaning From ECMO in Patients With Severe Cardiogenic Shock: The LEVOECMO Randomized Clinical Trial. *JAMA*. 2026;335(1):60-69.
7. Combes A, Saura O, Nessler N, Lebbah S, Rozec B, Levy B, et al. Supplement 3 to Levosimendan to Facilitate Weaning From ECMO in Patients With Severe Cardiogenic Shock: The LEVOECMO Randomized Clinical Trial. *JAMA*. 2025.
8. Paulo N, Kimmoun A, Hajage D, Hubert P, Levy D, Pineton de Chambrun M, et al. Does Levosimendan hasten veno-arterial ECMO weaning? A propensity score matching analysis. *Ann Intensive Care*. 2025;15(1):48.

EVERDAC

Muller G, Contou D, Ehrmann S, Martin M, Andreu P, Kamel T, et al. Deferring Arterial Catheterization in Critically Ill Patients with Shock. N Engl J Med. 2025;393(19):1875-1888.

Introduction

An arterial catheter is an established feature of shock management, providing continuous pressure measurement and convenient access for blood sampling. Its apparent advantages are immediate: beat-to-beat changes are visible, vasopressors can be titrated against a continuous trace and repeated arterial punctures may be avoided. These benefits in measurement and workflow do not necessarily translate into better patient outcomes.

Catheter insertion and continued use also create risks. Bleeding, haematoma, thrombosis and infection are possible, and easy access may encourage more frequent blood sampling. An automated arm cuff avoids arterial access but measures intermittently, may fail or disagree with invasive measurements, and can cause discomfort during repeated inflation. The clinical comparison is therefore between complete monitoring strategies, including rescue procedures and blood-sampling practices, rather than between two isolated readings.

Before EVERDAC, outcome evidence came chiefly from observational comparisons, vulnerable to the fact that clinicians preferentially insert arterial catheters in sicker patients. Diagnostic studies could assess whether cuff measurements identify hypotension, but could not determine the net effects of using a cuff-based strategy throughout an episode of shock. EVERDAC tested whether deferring arterial catheterisation, with predefined reasons to escalate, preserved acceptable short-term survival compared with routine early insertion.^{1,2,3}

Synopsis

EVERDAC was an investigator-initiated, pragmatic, open-label non-inferiority trial conducted in nine French intensive care units. Recruitment ran from November 2018 to November 2022. The French Ministry of Health funded the study. Patients or representatives received information and gave documented oral consent under the applicable French framework, with later patient consent when capacity returned. Clinical care and outcome assessment were not masked.^{3,4}

Adults were eligible within 24 hours of intensive care admission if they had acute circulatory failure: persistent systolic pressure below 90 mmHg or mean pressure below 65 mmHg for more than fifteen minutes, or initiation of a vasopressor, together with at least one sign of tissue hypoperfusion. A functioning upper-arm cuff measurement was required. The study included several shock aetiologies rather than sepsis alone.

Exclusions included inability to obtain a non-invasive reading, inability to place a brachial cuff, extracorporeal membrane oxygenation, severe traumatic brain injury, body mass index above 40 and very high vasopressor requirements. The dose threshold was more than 2.5 micrograms/kg/min of noradrenaline tartrate plus adrenaline, using the formulations specified by the investigators. Noradrenaline tartrate was expressed at twice the equivalent noradrenaline-base dose; the quoted threshold must not be transferred uncritically to a different formulation.

A central concealed web system randomised patients equally, with variable blocks and stratification by centre, invasive ventilation and vasopressor dose. The non-invasive strategy avoided arterial catheter insertion through day 28 unless a predefined safety criterion arose. An existing arterial catheter was to be removed within one hour. Pressure was measured with an intermittent automated brachial cuff, at a frequency chosen by the clinical team. Continuous non-invasive technologies were not used.

Rescue criteria included failure to obtain cuff pressure or pulse oximetry, an essential arterial blood gas after five failed puncture attempts, a requirement for ECMO, escalation beyond the specified vasopressor threshold or high-risk surgery. A catheter inserted for surgery was to be removed within four hours of return to intensive care. The comparator required catheter insertion within four hours of allocation if one was not already present. Cuff monitoring was generally restricted to insertion or replacement periods or circumstances in which invasive monitoring was no longer useful.

In both groups, catheters could be removed when their use was considered futile or unnecessary, including sustained improvement with low vasopressor requirements and no hypoperfusion, or a decision for palliative care. Other treatment followed local practice and international guidance. Central venous sampling, when available, was encouraged to reduce arterial punctures. No participating centre used pulse-contour monitoring in this trial, and continuous vasopressor treatment involved noradrenaline or adrenaline rather than vasopressin or other agents.

The primary endpoint was all-cause mortality at 28 days. The non-inferiority margin allowed at most five percentage points of excess mortality with the non-invasive strategy. Investigators anticipated mortality of 25% with invasive monitoring and 22.5%

with non-invasive monitoring, reflecting a possible small benefit from avoiding catheter-related harms. Under those assumptions, 1,010 patients would provide 80% power. A superiority analysis was planned if non-inferiority was established.

The primary model estimated an adjusted risk difference using a binomial generalised estimating equation with centre and randomisation strata. Intention-to-treat and two per-protocol analyses were reported. Secondary outcomes included 90-day mortality, organ failure, organ support, catheter-related infection, puncture burden, transfusion and device-related pain. Their analyses were exploratory without multiplicity adjustment. No subgroup analyses were prespecified.

The screening framework identified 4,183 admissions with potential circulatory failure, including 451 patients missed during prospective screening. A total of 1,010 were randomised, 506 to the non-invasive strategy and 504 to early invasive monitoring. One consent withdrawal and one guardianship-related exclusion in each arm left 504 and 502 for analysis. Mean age was 66–67 years, approximately two-thirds were men and mean Sequential Organ Failure Assessment score was ten. About 54% had septic shock, with smaller cardiogenic, haemorrhagic, post-arrest and other groups. Approximately 94% were medical admissions, two-thirds were invasively ventilated and 89% were receiving vasopressors at allocation.

Randomisation occurred about five and a half hours after intensive care admission. Sixty-four non-invasive patients already had an arterial catheter. Thirty-one had it removed within one hour, another 31 within 24 hours, one later and one not at all. During follow-up, 74 non-invasive patients, 14.7%, underwent catheterisation at a median of 22 hours; 68 did so because a safety criterion was met. Early invasive monitoring produced catheter use in 493 of 502 patients, 98.2%, including those with a catheter at enrolment. New insertion occurred a median of one hour after allocation.

At day 28, 173 of 504 non-invasive patients, 34.3%, and 185 of 502 invasive patients, 36.9%, had died. The adjusted risk difference was –3.2 percentage points, with a 95% confidence interval of –8.9 to 2.5. The upper limit lay below the five-point margin, establishing non-inferiority, with $P=0.006$. Superiority was not established, with $P=0.20$. The two per-protocol estimates had upper confidence limits of 2.6 percentage points and also met the non-inferiority criterion.^{3,4}

At 90 days, mortality was 42.7% versus 44.0%, with an adjusted risk difference of –1.7 percentage points and a 95% confidence interval of –7.0 to 3.5. Organ failure improved similarly and there was no clear difference in renal replacement or ventilator support-free days. Arterial-site haematoma or haemorrhage occurred in five versus 41

patients, 1.0% versus 8.2%. Arterial catheter-related bloodstream infection comprised three events in three non-invasive patients and fourteen events in thirteen invasive patients; the reported incidence ratio was 0.18, with a 95% confidence interval of 0.06–0.54.

The trade-off was more repeated arterial puncture for sampling. Excluding attempts to place catheters, the rates were 742 versus 269 attempts per 1,000 intensive care days, an incidence ratio of 2.76 with a 95% confidence interval of 2.41–3.16. Device-related pain or discomfort on at least one day was reported by 66 versus 45 patients. Among those able to answer, this represented 22.8% versus 15.4%. The question concerned the device in place and deliberately excluded pain during or immediately after a vascular puncture.

Critique

EVERDAC challenges automatic arterial catheterisation for every patient with shock. Its primary result supports a strategy of initial cuff monitoring with active reassessment and specified escalation criteria. It does not show that an arterial catheter is never useful, that cuff measurements are perfectly accurate or that non-invasive monitoring improves survival.

The non-inferiority margin is the first issue to examine. A margin of five percentage points is clinically substantial when the outcome is death. At the anticipated 25% control mortality, it would permit mortality as high as 30% before the strategy failed the margin. The observed data are more reassuring than that maximum allowance: the upper confidence limit was 2.5 percentage points, while the point estimate favoured non-invasive monitoring. Nevertheless, the study does not exclude every small increase in mortality, and a clinician's willingness to accept that uncertainty depends on the benefits and burdens of avoiding arterial access.

The sample calculation assumed a 2.5-point advantage for the non-invasive strategy, rather than identical mortality in the two groups. This matters because the planned distance from the non-inferiority boundary was larger than under an equal-outcome assumption. Actual mortality was also higher than anticipated. The final confidence interval, rather than the nominal power alone, is the useful summary: it excluded the chosen margin but still encompassed both benefit and a smaller degree of harm.

Non-inferiority designs are vulnerable when poor adherence makes treatment groups look alike. Here, however, the separation in catheter exposure was large, and the per-protocol estimates agreed closely with the principal result. Rescue catheterisation in the non-invasive arm is not automatically contamination: it was an intended component

of the strategy. The clinically relevant question is whether routine early insertion adds benefit when a catheter can still be placed for deterioration or practical failure.

The existing-catheter subgroup adds nuance. About 13% already had invasive monitoring at randomisation, and only half of those allocated to removal met the one-hour target. Some therefore received several additional hours of invasive monitoring during early shock. The study is not a pure comparison of never having an arterial catheter with always having one. It is a comparison of active deferral or removal against routine continued or early use, with a realistic degree of delay.

The per-protocol reporting itself has an unresolved detail. The supplementary definition refers broadly to excluding non-invasive patients catheterised within 24 hours, yet only two were excluded despite rescue insertions occurring during that period. The main flow appears to preserve rescue permitted by the strategy, which is clinically reasonable. The exact wording should be clarified rather than treating the labelled analyses as a separate proof that rescue had no effect. Their numerical agreement is reassuring but does not remove the need to understand what adherence meant.

This was a seriously ill cohort, not merely stable patients with a transient low reading. Mortality exceeded one-third, most required vasopressors and two-thirds were ventilated. That strengthens the relevance of testing whether invasive monitoring is indispensable. However, the requirement for an obtainable cuff pressure selected out patients in whom the proposed alternative already failed. Very high vasopressor doses, ECMO, severe brain injury and marked obesity were also exclusions. The result cannot authorise the same policy in those circumstances without further evidence.

Further selection occurred before randomisation. Some patients were missed, others excluded for logistical reasons or at physician discretion without a recorded explanation. This primarily limits representativeness; it does not by itself invalidate the randomised comparison among those enrolled. The predominance of medical admissions also leaves less certainty for rapidly changing perioperative haemodynamics, major trauma or situations requiring tightly controlled cerebral perfusion.

The trial assesses the net effect of monitoring strategies, not measurement equivalence. Neither accuracy against a reference standard nor the full frequency and duration of hypotension was the primary question. Continuous invasive data may be physiologically informative without improving survival under ordinary treatment practices. Conversely, a reassuring average outcome cannot guarantee that every individual cuff reading is adequate. Implausible readings, clinical deterioration and failure to obtain a value remain reasons for reassessment.

Reduced catheter-site bleeding and infection are plausible benefits of lower exposure, but patient comfort did not uniformly improve. Cuff-related discomfort was more frequent, and the non-invasive group experienced substantially more arterial punctures. The pain analysis excluded puncture-associated pain and was restricted to patients able to communicate. It therefore cannot estimate total procedural suffering across the whole cohort. Avoiding an indwelling catheter may exchange one burden for another unless blood-sampling practice also changes.

The infection denominator also needs attention. Rates were expressed per intensive care days, not per catheter days. They estimate the infection burden of the strategy, which is clinically useful, rather than the intrinsic infection risk of a catheter once inserted. Similarly, fewer arterial-site haematomas should not be interpreted as proof that insertion technique improved: fewer patients were exposed to insertion.

The unblinded design matters more for discretionary secondary outcomes than for the primary mortality endpoint. Clinicians determined sampling frequency, vasopressor withdrawal and when a catheter was no longer useful. Staff workload and satisfaction were not measured, although they could strongly influence implementation. The practical alternative is therefore not simply to remove the arterial line and continue exactly the same sampling habits; it requires a workable plan for observation, blood tests and escalation.

EVERDAC supports selective rather than reflexive catheterisation in a substantial group of patients with shock. The benefit is the opportunity to avoid an invasive device without exceeding the trial's mortality margin, not proof that non-invasive care is universally safer or more comfortable. Retaining access to prompt arterial monitoring when the alternative fails is integral to that conclusion.

Body of Evidence

A large observational study across 139 United States intensive care units examined arterial catheter use in mechanically ventilated medical patients. In 13,603 propensity-matched pairs, hospital mortality was similar, with an odds ratio of 0.98 and a 95% confidence interval of 0.93–1.03. A vasopressor subgroup showed an association with higher mortality, but residual differences in severity and treatment selection prevent a causal interpretation. These data questioned an assumed survival benefit without establishing that withholding the device was safe. EVERDAC supplies the randomised strategy comparison that such analyses could not provide.¹

A separate propensity analysis studied ventilated patients without sepsis or vasopressor requirements. Mortality was again similar after matching, but that population was haemodynamically stable. It was useful evidence against indiscriminate use solely because a patient was ventilated, while offering only indirect guidance for shock. EVERDAC extends the question to a substantially sicker cohort while retaining explicit limits on when cuff monitoring is acceptable.⁵

Measurement studies address a different question. In a study of 150 critically ill patients, including 83 with acute circulatory failure, brachial cuff mean pressure differed from invasive pressure by an average of 3.4 mmHg, with limits of agreement extending from approximately -6 to +13 mmHg. Detection of mean pressure below 65 mmHg was good at the group level. Nevertheless, a useful diagnostic test is not an interchangeable measurement in every patient. EVERDAC's strategy therefore depends on both the cuff reading and clinical assessment, with escalation when the non-invasive approach fails.²

Arterial access also has a measurable infectious burden. A prospective multicentre cohort comparing arterial and central venous catheters found similar catheter-related infection rates, approximately one event per 1,000 catheter days. This did not compare an arterial catheter with no catheter and cannot establish the effect of a deferral policy. It does, however, counter the assumption that arterial lines are effectively exempt from the infectious risks of intravascular devices.⁶

The broader history of invasive monitoring offers a useful caution. PAC-Man randomised critically ill patients to management with or without a pulmonary artery catheter and found no hospital survival advantage. That is not direct evidence about arterial pressure monitoring: the device, information and risks differ. Its relevance is the recurring distinction between obtaining more physiological information and demonstrating that the resulting management improves patient outcomes.⁷

The accompanying editorial places EVERDAC within a move towards examining whether familiar invasive practices are necessary for every patient. The trial supports that reassessment, but its practical alternative includes reliable cuff measurement, repeated clinical observation and ready access to rescue catheterisation. A service adopting deferral should also reconsider routine arterial blood-gas sampling, since otherwise the reduction in indwelling devices may be accompanied by many more punctures.⁸

Can we defer an arterial catheter in patients with shock?

Yes, in patients resembling those enrolled in EVERDAC, reliable cuff monitoring with explicit rescue criteria was non-inferior for 28-day mortality. The result supports selective catheterisation, with prompt escalation when non-invasive monitoring is inadequate.

References

1. Gershengorn HB, Wunsch H, Scales DC, Zarychanski R, Rubenfeld G, Garland A. Association between arterial catheter use and hospital mortality in intensive care units. *JAMA Intern Med.* 2014;174(11):1746-1754.
2. Lakhal K, Macq C, Ehrmann S, Boulain T, Capdevila X. Noninvasive monitoring of blood pressure in the critically ill: reliability according to the cuff site (arm, thigh, or ankle). *Crit Care Med.* 2012;40(4):1207-1213.
3. Muller G, Contou D, Ehrmann S, Martin M, Andreu P, Kamel T, et al. Deferring Arterial Catheterization in Critically Ill Patients with Shock. *N Engl J Med.* 2025;393(19):1875-1888.
4. Muller G, et al. Supplementary Appendix to Deferring Arterial Catheterization in Critically Ill Patients with Shock. *N Engl J Med.* 2025.
5. Hsu DJ, Feng M, Kothari R, Zhou H, Chen KP, Celi LA. The Association Between Indwelling Arterial Catheters and Mortality in Hemodynamically Stable Patients With Respiratory Failure: A Propensity Score Analysis. *Chest.* 2015;148(6):1470-1476.
6. Lucet JC, Bouadma L, Zahar JR, Schwebel C, Geffroy A, Pease S, et al. Infectious risk associated with arterial catheters compared with central venous catheters. *Crit Care Med.* 2010;38(4):1030-1035.
7. Harvey S, Harrison DA, Singer M, Ashcroft J, Jones CM, Elbourne D, et al. Assessment of the clinical effectiveness of pulmonary artery catheters in management of patients in intensive care (PAC-Man): a randomised controlled trial. *Lancet.* 2005;366(9484):472-477.
8. Gershengorn HB. A Less Invasive Approach to Intensive Care. *N Engl J Med.* 2025;393(19):1953-1954.

FLUID

McIntyre L, Fergusson D, McArdle T, English S, Cook DJ, Fox-Robichaud AE, et al. A Crossover Trial of Hospital-Wide Lactated Ringer's Solution versus Normal Saline. N Engl J Med. 2025;393(7):660-670.

Introduction

Crystalloid selection is a small decision repeated on an enormous scale. Saline is familiar and widely available, but its chloride concentration exceeds that of plasma. Balanced solutions reduce chloride exposure and contain anions that act as buffers. These physiological differences provide a rationale for benefit, particularly when large volumes are administered, but do not themselves establish that changing fluid improves survival or prevents clinically important kidney injury.

Most major comparative trials have concentrated on critically ill patients or on fluid given in the emergency department. Hospitals, however, must also decide which fluid to stock as the default across wards, theatres and assessment areas. A modest effect that is difficult to detect in an individual patient could matter across a whole health system. Conversely, many hospitalised patients receive little or no crystalloid, potentially diluting the effect of a hospital-wide policy.

FLUID addressed this organisational question. It compared policies favouring lactated Ringer's solution or normal saline throughout participating hospitals, using routinely collected data to measure death and readmission. Its scale is impressive, but its unit of randomisation and the quantity it estimates are central to interpreting the result: this was a trial in seven hospital clusters, not 43,626 independent individual allocations.¹

Synopsis

FLUID was an open-label, pragmatic, cluster-randomised crossover trial conducted in seven academic and community hospitals in Ontario, Canada, between August 2016 and March 2020. Funding came from the Canadian Institutes of Health Research and the Ottawa Hospital Academic Medical Organization. Baxter Healthcare supported fluid availability without a role in trial design, analysis or reporting. Ethics committees approved a waiver of individual consent because the intervention was a hospital-wide substitution between usual treatments.^{1,2}

Hospitals required level II or III intensive care capacity and at least 1,500 admissions during each trial period. Eligible patients had an index admission during an active period, defined as a first admission without another hospital admission in the preceding 90 days. Patients were included regardless of whether they actually received intravenous fluid. Infants younger than one month, patients with a missing birth date and admissions during run-in or washout periods were excluded, alongside administrative-data eligibility and quality exclusions.

An independent statistician generated the hospital allocation sequence in blocks of two. Three hospitals used lactated Ringer's solution first and saline second; four followed the opposite sequence. Sites were informed approximately one month before starting to allow preparation. Each hospital completed two active twelve-week periods. Before each, one to three weeks were allowed to stock the assigned fluid, followed by a one-week run-in. After each active period, a one-week washout at the first five sites, or two weeks at the final two, allowed patients admitted near the end to continue receiving the assigned fluid. Newly admitted patients during these transition periods did not enter the analysis.

The assigned fluid was intended to constitute at least 80% of crystalloid inventory throughout each hospital. Electronic systems automatically substituted it when a clinician ordered saline or lactated Ringer's solution; nurses implemented substitution where orders were manual. Physicians could override the choice for clinical reasons. Such patients remained in their assigned hospital-period analysis. Inventory reports monitored fluid use, but individual receipt, timing and volume were unavailable.

The primary outcome was a composite of death or hospital readmission within 90 days after the index admission. Secondary outcomes included the individual components, emergency department attendance, initiation of dialysis, hospital length of stay and discharge somewhere other than home. Linked provincial administrative databases supplied all outcomes. Serious adverse events considered related to study fluid were reported through existing hospital safety processes.

The primary estimand was the average effect across hospitals. Each hospital received equal weight, rather than weighting larger hospitals more heavily because they contributed more patients. Outcome proportions were calculated for each hospital-period and analysed using regression terms for treatment, hospital and first versus second period. Prespecified adjustment incorporated age, sex, case mix, comorbidity and admission to intensive care within one day. A two-stage approach first estimated expected outcomes from patient characteristics and then compared

hospital-period residuals. Confidence intervals used a t distribution with degrees of freedom equal to the number of hospitals minus two.

The initial calculation required twelve hospitals for 80% power to detect a reduction in the primary outcome from 16% to 15%. Four additional hospitals were planned to accommodate non-adherence and variation in cluster size, giving a target of sixteen hospitals and approximately 144,000 admissions. The COVID-19 pandemic made further conduct impracticable, and recruitment stopped after seven hospitals had completed both periods. This was an external operational interruption rather than stopping for demonstrated benefit or futility.

The analysis included 43,626 patients: 22,017 during lactated Ringer's periods and 21,609 during saline periods. Mean age was approximately 59 years and 58% were female. Medical and surgical admissions each accounted for about two-fifths of the population, pregnancy and childbirth for 17%, and mental health admissions for less than 1%. Approximately 10% entered intensive care within one day. Baseline characteristics were broadly similar. No primary or secondary outcome data were missing.

Inventory-based adherence was 78.2% during lactated Ringer's periods and 93.6% during saline periods. The average hospital incidence of death or readmission was 20.3% versus 21.4%. After adjustment, the absolute difference was -0.53 percentage points, with a 95% confidence interval of -1.85 to 0.79; $P=0.35$. The adjusted relative risk was 0.97, with a 95% confidence interval of 0.90–1.05. These are hospital-average estimates, not pooled individual-patient risks.^{1,2}

The average hospital mortality proportions were 6.9% and 7.6%, with an adjusted difference of -0.26 percentage points and a 95% confidence interval of -0.95 to 0.43. Readmission proportions were 15.1% and 15.4%, with an adjusted difference of -0.06 points and a 95% confidence interval of -1.78 to 1.67. Dialysis initiation was uncommon, averaging 0.5% and 0.6%. No secondary outcome demonstrated a clear difference. The intervals were not adjusted for multiple comparisons.

Sensitivity analyses accounting for calendar time, excluding selected lower-risk admission categories or removing patients who might span fluid periods produced similar primary estimates. Subgroup analyses were also broadly compatible with the overall result, with considerable uncertainty in small clinical subgroups. Participating sites reported no serious adverse events to the central coordinating centre.

Critique

FLUID did not demonstrate that a hospital-wide lactated Ringer's policy reduced death or readmission. It also did not establish that the two policies are equivalent. The confidence interval permits an absolute reduction approaching two percentage points or an increase of less than one point in the average hospital. A one-point benefit, the effect considered worthwhile when planning the trial, remains compatible with the data.

The distinction between the large patient count and the small number of randomised hospitals is fundamental. Thousands of admissions improve the precision of each hospital's event rate, but do not create thousands of independently allocated treatment contrasts. With seven clusters, the primary inference used only five residual degrees of freedom. The crossover design helps because each hospital contributes experience under both policies, but it cannot fully replace the planned number of hospitals. Describing the study simply as a very large negative trial obscures this limitation.

Equal weighting also determines whose question is answered. FLUID estimates the effect for an average participating hospital, giving a smaller hospital the same influence as a larger one. This is defensible for a policy decision across institutions, but differs from the average effect across all individual patients. The reported 20.3% and 21.4% are means of hospital proportions, not event totals divided by the two patient denominators. Likewise, their arithmetic difference is not the adjusted treatment effect of -0.53 percentage points.

The intervention is a default policy with clinical exceptions. It does not compare known quantities of saline and lactated Ringer's solution in patients confirmed to require fluid. Inclusion of every eligible admission makes the policy evaluation pragmatic and avoids selecting patients according to a treatment decision made after allocation. The cost is uncertainty about the actual biological contrast. A person who received no crystalloid contributes the same type of outcome information as someone who received several litres, although only the latter had substantial exposure to the fluid difference.

This dilution does not make the study irrelevant. If the intended decision is to change the default throughout a hospital, outcomes among all admissions are relevant. It does limit translation to an individual with shock, where volume, timing, baseline chloride, renal function and prior resuscitation may matter. FLUID cannot establish whether a larger effect exists among patients receiving high volumes, because their exposure was not recorded.

Inventory adherence is another important distinction. It reflects the proportion of fluid used across a hospital, not the proportion of individual patients who received their

assigned treatment exclusively. The lower adherence to lactated Ringer's solution, just below the 80% target, reduces separation between policies. It may reflect familiarity, indications for overriding the default or practical implementation. Without patient-level reasons or doses, the study cannot distinguish these explanations or construct a reliable exposure-response analysis.

Each hospital crossed over once. This controls stable differences between institutions but leaves vulnerability to time-varying influences, including seasonal case mix and changes in clinical practice. Sites began at different calendar times, so a simple first-versus-second-period term does not fully represent the same season or year across hospitals. Additional models using linear calendar time or a restricted cubic spline were reassuringly similar, although a small number of clusters limits how completely complex temporal effects can be modelled.

The transition periods were sensible attempts to reduce exposure to both fluids during one admission. Nevertheless, a long stay could bridge a switch, and an admission near the end of the second period could continue after usual practice resumed. Excluding patients who might experience such overlap did not materially alter the estimate. That sensitivity analysis strengthens the robustness of the policy result without showing that every remaining patient received a pure fluid exposure.

The population is much broader and generally less acutely ill than those in ICU fluid trials. Obstetric admissions were common, and only one in ten patients entered intensive care promptly. This is a strength for a hospital-wide question and a limit for a critical care decision. The ICU subgroup contained 4,372 patients, but subgroup inference still depended on the seven hospitals and was imprecise. The small sepsis and traumatic brain injury subgroups cannot overturn more substantial evidence in those conditions.

Several subgroup point estimates favoured lactated Ringer's solution, and the interval in patients older than 80 years excluded zero. These analyses involved many comparisons, lacked multiplicity adjustment and were not the primary test. A within-subgroup interval also does not establish that treatment effects differ between age groups. The study provides no secure basis for an age-specific fluid policy.

Administrative data made a broad trial feasible and achieved complete outcome ascertainment. Death and readmission are meaningful outcomes, and their capture was reported to be highly accurate. However, codes provide less clinical detail than prospectively collected physiology. Dialysis is a severe and relatively rare kidney outcome; its absence does not establish the absence of acute kidney injury. Electrolyte

changes, acid–base effects and creatinine trajectories were not available to explain how the two policies affected patients.

There is also a reporting issue concerning follow-up timing. The main report defines the primary composite from the index admission, whereas parts of the supplementary data dictionary describe readmission and several secondary outcomes using time after discharge. These formulations are not interchangeable for patients with long stays. The chapter retains the primary definition in the main report, but the exact implementation should be clarified before reproducing the analysis. This discrepancy does not supply evidence for a different treatment result; it limits certainty about the outcome's precise construction.²

The main outcome table's adjusted length-of-stay estimate is also difficult to reconcile with its stated units and the displayed hospital means. It is not used here to infer a clinically meaningful duration effect. Accurate interpretation is better served by leaving an unclear estimate unresolved than by silently converting or correcting it.

The supplementary account explains a delay in public registration of the larger trial, while stating that the clinical outcomes had appeared in the pilot programme and that outcome data had not been examined before registration. This explanation is relevant, but delayed public availability weakens independent verification of prospective decisions. The present review used the published report and supplementary methods and results, without using separate protocol or statistical analysis plan documents.

Finally, absence of reported serious adverse events should not be interpreted as proof of identical safety. Surveillance relied on ordinary safety meetings and reporting of events considered fluid-related, rather than systematic collection of every biochemical or clinical complication. The broad outcome data are reassuring against a conspicuous population-level difference, but they cannot exclude uncommon harms or effects in specialised groups.

FLUID therefore leaves a bounded conclusion: changing the hospital default did not produce a demonstrable improvement in its composite outcome, but modest benefit remains possible. It adds evidence about implementation across a health system while leaving the choice for high-risk, fluid-exposed patients dependent on the wider evidence and their clinical circumstances.

Body of Evidence

SALT-ED studied 13,347 adults receiving crystalloid in an emergency department before admission outside intensive care. Hospital-free days, the primary outcome, were similar. Major adverse kidney events, a secondary composite, occurred in 4.7% with balanced crystalloids and 5.6% with saline. Unlike FLUID, these patients were known to have received fluid, and allocation concerned emergency department treatment rather than a whole-hospital policy.³

SMART enrolled 15,802 adults in five intensive care units using a cluster crossover design. The primary composite of death, new renal replacement or persistent renal dysfunction occurred in 14.3% with balanced crystalloids and 15.4% with saline, with a marginal odds ratio of 0.91 and a 95% confidence interval of 0.84–0.99. Mortality alone was not clearly different. The modest composite benefit helped motivate wider interest in balanced fluids, but its population, endpoint and exposure differ from FLUID.⁴

The individually randomised, masked BaSICS trial analysed 10,520 patients in Brazilian intensive care units. Ninety-day mortality was 26.4% with a balanced solution and 27.2% with saline, with an adjusted hazard ratio of 0.97 and a 95% confidence interval of 0.90–1.05. PLUS recruited 5,037 patients in Australia and New Zealand and likewise found similar 90-day mortality, 21.8% versus 22.0%, without a clear difference in new renal replacement. Both limit expectations of a large general survival effect.^{5,6}

The BEST-Living individual-patient meta-analysis combined six ICU trials involving 34,685 participants. Its Bayesian estimate suggested a small possible hospital mortality benefit with balanced solutions: an odds ratio of 0.962, with a 95% credible interval of 0.909–1.019 and an 89.5% probability of benefit. The traumatic brain injury subgroup instead raised concern for harm. These findings favour a nuanced interpretation rather than treating all balanced fluids and all patient groups as interchangeable.⁷

FLUID is compatible with a small overall benefit but does not confirm one. It concerns lactated Ringer's solution specifically, and its unmeasured individual exposure and hospital-average analysis should be retained when comparing it with trials of other balanced solutions. Its contribution is a pragmatic test of changing the institutional default, not a final resolution of crystalloid choice in every critically ill patient.

Should FLUID change our choice of crystalloid?

FLUID does not by itself justify changing practice: a hospital-wide lactated Ringer's policy did not significantly reduce death or readmission, although modest benefit remains possible. Individual fluid selection should still reflect the wider evidence and the patient's clinical context.

References

1. McIntyre L, Fergusson D, McArdle T, English S, Cook DJ, Fox-Robichaud AE, et al. A Crossover Trial of Hospital-Wide Lactated Ringer's Solution versus Normal Saline. *N Engl J Med*. 2025;393(7):660-670.
2. McIntyre L, et al. Supplementary Appendix to A Crossover Trial of Hospital-Wide Lactated Ringer's Solution versus Normal Saline. *N Engl J Med*. 2025.
3. Self WH, Semler MW, Wanderer JP, Wang L, Byrne DW, Collins SP, et al. Balanced Crystalloids versus Saline in Noncritically Ill Adults. *N Engl J Med*. 2018;378(9):819-828.
4. Semler MW, Self WH, Wanderer JP, Ehrenfeld JM, Wang L, Byrne DW, et al. Balanced Crystalloids versus Saline in Critically Ill Adults. *N Engl J Med*. 2018;378(9):829-839.
5. Zampieri FG, Machado FR, Biondi RS, Freitas FGR, Veiga VC, Figueiredo RC, et al. Effect of Intravenous Fluid Treatment With a Balanced Solution vs 0.9% Saline Solution on Mortality in Critically Ill Patients: The BaSICS Randomized Clinical Trial. *JAMA*. 2021;326(9):818. doi:10.1001/jama.2021.11684.
6. Finfer S, Micallef S, Hammond N, Navarra L, Bellomo R, Billot L, et al. Balanced Multielectrolyte Solution versus Saline in Critically Ill Adults. *N Engl J Med*. 2022;386(9):815-826.
7. Zampieri FG, Cavalcanti AB, Di Tanna GL, Damiani LP, Hammond NE, Machado FR, et al. Balanced crystalloids versus saline for critically ill patients (BEST-Living): a systematic review and individual patient data meta-analysis. *Lancet Respir Med*. 2024;12(3):237-246.

STORM-PE

Lookstein RA, Konstantinides SV, Weinberg I, Dohad SY, Rosol Z, Kopeć G, et al. Randomized Controlled Trial of Mechanical Thrombectomy With Anticoagulation Versus Anticoagulation Alone for Acute Intermediate-High Risk Pulmonary Embolism: Primary Outcomes From the STORM-PE Trial. Circulation. 2026;153(1):21-34. First published online 3 November 2025.

Introduction

Patients with acute pulmonary embolism can remain normotensive despite substantial right ventricular strain. Some recover with anticoagulation and supportive care, while others deteriorate as the right ventricle fails. Identifying those who benefit from immediate reperfusion, before haemodynamic collapse occurs, remains difficult. The potential advantage of removing thrombus must be weighed against bleeding, procedural injury and treatment of patients who would otherwise improve without an intervention.

Systemic thrombolysis can prevent some episodes of deterioration but carries an important bleeding cost. Catheter-based treatment offers alternative ways to reduce obstruction, either with local thrombolysis or mechanical extraction. Mechanical aspiration avoids administering a thrombolytic drug, although it introduces its own risks and does not remove the need for anticoagulation. Evidence of rapid physiological improvement has often come from uncontrolled device studies, in which improvement with time and anticoagulation cannot be separated from the intervention's effect.

STORM-PE provides a randomised comparison of computer-assisted vacuum thrombectomy plus anticoagulation against anticoagulation alone. Its primary question was whether thrombectomy produced a greater reduction in the right-to-left ventricular diameter ratio at 48 hours. That is an important step in evaluating the device, but it is a surrogate question: better early imaging is not automatically proof of better survival, recovery or long-term function.¹

Synopsis

STORM-PE was an open-label, international, post-market randomised trial at 22 sites in the United States, Canada, New Zealand and Poland. Recruitment occurred between July 2023 and June 2025. Penumbra, the manufacturer, sponsored and funded the trial and participated with the investigators in design, data collection, analysis, interpretation and publication. An independent imaging core laboratory, masked to treatment allocation, assessed the primary endpoint, and an external clinical events committee adjudicated safety events.^{1,2}

Eligible adults were aged 18–80 years, had symptoms for no more than fourteen days and were normotensive with acute pulmonary embolism. CT pulmonary angiography had to show a right-to-left ventricular diameter ratio of at least 1.0, with an elevated cardiac biomarker: troponin, BNP or NT-proBNP. Key exclusions included haemodynamic instability, a National Early Warning Score 2 of at least 9 and a contraindication to anticoagulation. The trial therefore selected patients with right ventricular strain while excluding those with overt instability or particularly high clinical deterioration scores.

An online system allocated treatment equally using fixed permuted blocks, stratified by centre. All patients received unfractionated or low-molecular-weight heparin according to clinical judgement, including treatment started before randomisation. Unfractionated heparin was adjusted to therapeutic laboratory targets and low-molecular-weight heparin to weight and renal function. Subsequent anticoagulant selection was discretionary, with at least three months of treatment expected.

The intervention added aspiration thrombectomy using the 16F Lightning Flash system. Its computer-controlled aspiration was intended to remove thrombus while limiting blood loss. Treatment was to begin within twelve hours of randomisation and 24 hours of the diagnostic CT. Venous access was obtained, the catheter was advanced through the right heart into the pulmonary arteries, and aspiration continued according to the operator's judgement of thrombus removal and haemodynamic or clinical improvement. No thrombolytic drug was routinely part of the index procedure.

The primary endpoint was the reduction in right-to-left ventricular diameter ratio between baseline and CT performed 48 hours after randomisation, with a six-hour window. The core laboratory measured maximal ventricular diameters on axial images. The sample calculation assumed a ratio reduction of 0.35 with thrombectomy and 0.10 with anticoagulation, giving an expected between-group difference of 0.25. One hundred randomised participants were intended to provide approximately 90% power for this imaging comparison.

For patients requiring rescue treatment without a preceding repeat ventricular measurement, the analysis assigned zero improvement in the ratio. The principal secondary safety endpoint was a seven-day composite of clinical deterioration requiring rescue, pulmonary embolism-related death, symptomatic recurrent embolism and major bleeding. Major bleeding used BARC categories 3a, 3b, 3c or 5, with a specified exception for an expected dilutional haemoglobin decrease and transfusion of fewer than two units. The primary report also examined vital signs, obstruction scores and hospital resource use. Longer-term function and quality of life were reserved for subsequent reporting.

Of 767 patients initially assessed, 100 were randomised, 47 to thrombectomy and 53 to anticoagulation alone. Eighty-six eligible patients declined participation. Mean age was approximately 60 years, and women comprised 38% versus 53% of the two groups. Baseline ventricular ratios were 1.63 and 1.56. Hypertension was more common in the anticoagulation group, as was the combination of elevated troponin and a natriuretic peptide, although all participants met the biomarker eligibility criterion.

All 47 assigned thrombectomy procedures were technically successful, defined as reaching the target thrombus and performing aspiration. All used femoral access. Median device time was 25 minutes and total procedure time 56 minutes. Mean estimated blood loss was approximately 297 mL. No adjunctive thrombolysis was used during the index procedures. Mean pulmonary artery pressure fell during treatment, but this intraprocedural comparison had no corresponding invasive control measurement.

Therapeutic anticoagulation was reached within 48 hours in 40 of 46 evaluable thrombectomy patients and 47 of 52 controls, 87% and 90%. Among patients receiving unfractionated heparin who reached a therapeutic laboratory value, median time from starting treatment was 6.8 versus 6.4 hours. These findings do not suggest an obvious group difference in the initial delivery of anticoagulation, although they do not establish uninterrupted therapeutic exposure in every patient.

The primary analysis included 46 thrombectomy and 52 control patients. One patient per group lacked the required CT assessment, and one per group received the assigned zero-change value after rescue without prior reassessment. Mean reduction in the ventricular ratio was 0.52 versus 0.24, with a between-group difference of 0.27 and a 95% confidence interval of 0.12–0.43; $P < 0.001$. The 48-hour ratios were 1.11 and 1.32. Thus, both groups improved, with a greater early improvement after thrombectomy.¹

Pulmonary artery obstruction scores also improved more with thrombectomy. An exploratory analysis found a ratio of 1.0 or less at 48 hours in 39.1% versus 13.5%.

Oxygen supplementation fell more in the intervention group. Heart rate and NEWS2 were lower at 48 hours, although the between-group differences in their changes from baseline were not statistically significant. Median hospital stay was five days in each group, without a clear difference in ICU admission or ICU length of stay.

The seven-day major adverse event composite occurred in two of 47 thrombectomy patients, 4.3%, and four of 53 controls, 7.5%. The risk difference was -3.3 percentage points, with a 95% confidence interval of -14.6 to 7.9; $P=0.681$. Two pulmonary embolism-related deaths occurred in the thrombectomy group and none in the control group. Clinical deterioration requiring rescue occurred in one versus three patients, and major bleeding in one patient per group. No symptomatic recurrent embolism was reported.

The first thrombectomy-group death followed massive haemoptysis and cardiac arrest after catheter removal during the procedure; autopsy found no cardiac or vascular injury. This patient contributed death, major bleeding and rescue events to the composite. The second death occurred five days later during a separate iliofemoral thrombectomy procedure. The events committee classified both deaths as pulmonary embolism-related and neither as related to the index thrombectomy procedure or device. These attribution decisions do not remove the deaths from the randomised comparison.²

Mean haemoglobin decline to the measured nadir was greater with thrombectomy, 2.5 versus 1.6 g/dL. No access-site complication was reported in the intervention group. Serious adverse events within seven days occurred in four versus eight patients. All these clinical comparisons were based on small event counts.

Critique

STORM-PE provides credible evidence that this aspiration system accelerates improvement in right ventricular size relative to anticoagulation alone. Randomisation addresses the principal weakness of earlier uncontrolled device studies, and masked core laboratory assessment reduces the risk that expectations influenced CT measurement. The positive primary result should be recognised for the question it answers.

The unresolved question is how much that imaging advantage benefits patients. Right ventricular enlargement is prognostic, but a prognostic marker is not automatically a validated surrogate for treatment benefit. A treatment can improve a marker while adding complications, and patients whose marker improves more rapidly may eventually

have similar recovery. The trial was not powered to demonstrate fewer deaths, less clinical deterioration or better long-term function.

The control group's improvement is itself instructive. Its mean ratio reduction was 0.24, more than twice the 0.10 assumed in planning. Anticoagulation does not mechanically remove clot, but endogenous resolution and recovery can still produce substantial improvement over 48 hours. Comparisons with historical device series cannot account reliably for this background recovery. The randomised between-group difference, rather than the intervention group's before-and-after change, is the relevant estimate of added efficacy.

Baseline ventricular ratio was somewhat higher in the intervention group, leaving more room for reduction. The primary analysis compared changes rather than a baseline-adjusted follow-up outcome. In a small trial, chance differences can affect precision and the apparent magnitude of change. The lower final ratio in the thrombectomy group supports a real early physiological effect, but it does not eliminate the need for larger trials with clinically important endpoints.

The primary analysis was described as intention-to-treat, yet the reported estimate used 98 of the 100 randomised participants. The missing assessments were numerically balanced but clinically different: one thrombectomy patient left hospital before repeat CT, while one control patient underwent surgical embolectomy. Missingness was therefore not simply a random technical failure, and the distinction should remain explicit. Assigning zero improvement after rescue is also a clinical convention, not an observed CT result. It prevents a successful rescue procedure from being counted as improvement under the originally assigned strategy, but it makes the endpoint partly dependent on the decision and timing of rescue.

One assigned zero in the intervention arm followed fatal deterioration. A zero change does not express the full clinical importance of death; it places that outcome on the same numerical scale as an unchanged ventricular ratio in a survivor. That limitation reinforces why the imaging result and safety outcomes must be read together. A positive mean imaging comparison cannot override a potentially important adverse event simply because the event received a neutral numerical change.

The safety composite is far too imprecise to establish equivalence. Its confidence interval allows a meaningful increase or decrease in adverse events. The observed two versus four affected patients should not be presented as proof of similar safety or as a reliable reduction in harm. Two deaths versus none is likewise insufficient to establish

mortality harm, but it is a clinically important observation that deserves visibility alongside the favourable imaging result.

Component overlap also matters. Three adverse outcomes occurred in one thrombectomy patient, while the composite counted that patient once. Conversely, three control patients deteriorated and received rescue treatment without dying. A composite that counts death, bleeding and rescue together can conceal very different clinical experiences. The total alone is therefore less informative than the component descriptions.

The attribution of the deaths warrants balanced treatment. The independent committee did not attribute either to the index device or procedure, and the absence of vascular injury at autopsy is relevant to the first event. Nevertheless, its temporal proximity to catheter removal and fatal pulmonary bleeding should be reported transparently. Causation cannot be independently resolved from the available account. Randomised safety assessment appropriately retains all events irrespective of whether a committee considers them device-related.

The supplement describes a further control patient who underwent emergency surgical embolectomy and removal of thrombus extending across a patent foramen ovale. This was considered rescue by the local investigator but did not satisfy the committee's endpoint classification for clinical deterioration requiring rescue. It therefore sits outside the three adjudicated rescue events. This is not necessarily incorrect, but it shows that the composite measures a defined adjudicated category rather than every escalation to an invasive treatment.²

Avoiding thrombolytic drugs does not mean avoiding bleeding. Aspiration removed an estimated average of almost 300 mL of blood, and the intervention group had a larger haemoglobin decline. A major bleeding definition captures severe events but does not fully describe blood loss, anaemia or the consequences for patients with limited reserve. The device's aspiration control may limit unnecessary loss, but STORM-PE does not compare that feature with another thrombectomy system.

The population was carefully selected. Haemodynamic instability and NEWS2 of at least 9 were exclusions, so the trial cannot establish outcomes for patients already approaching collapse. At the same time, patients without ventricular enlargement or biomarker elevation were excluded. Biomarker eligibility allowed troponin or natriuretic peptide elevation, rather than requiring both, which should be remembered when matching the trial to other risk classifications. Only a minority of assessed patients entered the study.

All sites had organised pulmonary embolism pathways and relevant procedural expertise. Twenty-two centres contribute breadth, but only 100 patients provide limited experience per centre. Technical success in this setting cannot be assumed in every hospital. The need for catheter facilities, trained operators, repeat imaging and management of complications remains part of the intervention's practical burden.

Open-label treatment particularly affects the softer outcomes. Oxygen withdrawal, ICU admission, discharge and rescue decisions may be influenced by knowledge of a successful procedure. Independent adjudication improves consistency in classifying events but does not blind the decisions that led to them. The absence of a hospital length-of-stay benefit is therefore relevant when considering suggestions that more rapid physiological improvement might improve patient throughput.

Many secondary and exploratory comparisons were tested without a clear multiplicity hierarchy protecting them. Lower heart rate or NEWS2 at 48 hours should not be confused with a demonstrated difference in their improvement from baseline. Similarly, a significant change within one group and a non-significant change within the other does not itself establish a between-group effect. The oxygen-flow comparison supports a possible early clinical benefit, but the broader secondary pattern is best treated as supportive rather than confirmatory.

The sponsor's extensive role does not invalidate the findings, and independent core laboratory and event committee assessments are meaningful safeguards. It does increase the importance of transparent reporting, retaining unfavourable events and avoiding claims beyond the prespecified imaging endpoint. The trial establishes efficacy for a particular system and setting; it does not prove superiority over competing catheter therapies or justify a class-wide conclusion about all mechanical thrombectomy.

Longer-term symptoms, exercise capacity and quality of life are central to whether an early imaging advantage is worth the procedure. Those outcomes were not available in this primary report. Within the book's 2025 evidence cutoff, the appropriate conclusion remains a positive physiological trial with unresolved net clinical benefit.

Body of Evidence

PEITHO randomised 1,006 normotensive patients with right ventricular dysfunction and positive troponin to tenecteplase or placebo alongside heparin. Death or haemodynamic decompensation within seven days occurred in 2.6% versus 5.6%, but major extracranial bleeding increased from 1.2% to 6.3%, with an increased stroke risk. Mortality alone was not clearly reduced. This established why preventing deterioration must be considered together with treatment-related harm when escalating reperfusion in intermediate-risk embolism.³

ULTIMA provided an earlier small randomised comparison of ultrasound-assisted catheter thrombolysis against heparin alone. Among 59 patients, local thrombolysis improved the ventricular ratio more rapidly at 24 hours. Like STORM-PE, it demonstrated a physiological effect but was too small to resolve major clinical outcomes. It tested a different intervention, and its results cannot be used to rank aspiration against catheter thrombolysis.⁴

EXTRACT-PE evaluated Indigo aspiration in 119 patients without a randomised control group and reported a mean ventricular ratio reduction of 0.43 at 48 hours. FLARE assessed a different mechanical system in 106 treated patients and likewise found an early reduction in right ventricular enlargement. These studies supported feasibility and further investigation, but their uncontrolled design could not distinguish device effects from recovery under anticoagulation. STORM-PE materially improves that evidence by supplying a contemporaneous randomised comparator.^{5,6}

PEERLESS randomised 550 patients to large-bore mechanical thrombectomy or catheter-directed thrombolysis. Its hierarchical composite favoured thrombectomy, driven by less deterioration or bailout and less postprocedural ICU use; mortality and major bleeding were not clearly different. Crucially, every participant was assigned a catheter intervention. It helps compare two invasive strategies in selected patients but cannot determine whether either is preferable to initial anticoagulation alone.⁷

These studies should not be arranged into a league table using their before-and-after ventricular changes. Devices, eligibility, timing, imaging methods and comparators differ. STORM-PE answers a previously underexamined comparison and demonstrates faster physiological recovery with this aspiration system. A larger trial measuring patient-centred benefit and harm is still needed before routine early thrombectomy can be justified for all patients with intermediate-high risk pulmonary embolism.

Should we routinely add mechanical thrombectomy for intermediate-high risk pulmonary embolism?

Not on the basis of STORM-PE alone. It improved early right ventricular imaging, but the trial was too small to establish a net benefit in survival, major complications or longer-term recovery.

References

1. Lookstein RA, Konstantinides SV, Weinberg I, Dohad SY, Rosol Z, Kopeć G, et al. Randomized Controlled Trial of Mechanical Thrombectomy With Anticoagulation Versus Anticoagulation Alone for Acute Intermediate-High Risk Pulmonary Embolism: Primary Outcomes From the STORM-PE Trial. *Circulation*. 2026;153(1):21-34.
2. Lookstein RA, et al. Supplemental Material to Primary Outcomes From the STORM-PE Trial. *Circulation*. 2026.
3. Meyer G, Vicaut E, Danays T, Agnelli G, Becattini C, Beyer-Westendorf J, et al. Fibrinolysis for patients with intermediate-risk pulmonary embolism. *N Engl J Med*. 2014;370(15):1402-1411.
4. Kucher N, Boekstegers P, Müller OJ, Kupatt C, Beyer-Westendorf J, Heitzer T, et al. Randomized, controlled trial of ultrasound-assisted catheter-directed thrombolysis for acute intermediate-risk pulmonary embolism. *Circulation*. 2014;129(4):479-486.
5. Sista AK, Horowitz JM, Tapson VF, Rosenberg M, Elder MD, Schiro BJ, et al. Indigo Aspiration System for Treatment of Pulmonary Embolism: Results of the EXTRACT-PE Trial. *JACC Cardiovasc Interv*. 2021;14(3):319-329.
6. Tu T, Toma C, Tapson VF, Adams C, Jaber WA, Silver M, et al. A Prospective, Single-Arm, Multicenter Trial of Catheter-Directed Mechanical Thrombectomy for Intermediate-Risk Acute Pulmonary Embolism: The FLARE Study. *JACC Cardiovasc Interv*. 2019;12(9):859-869.
7. Jaber WA, Gonsalves CF, Stortecky S, Horr S, Pappas O, Gandhi RT, et al. Large-Bore Mechanical Thrombectomy Versus Catheter-Directed Thrombolysis in the Management of Intermediate-Risk Pulmonary Embolism: Primary Results of the PEERLESS Randomized Controlled Trial. *Circulation*. 2025;151(5):260-273.

CELEBRATE

van't Hof AWJ, Gibson CM, Rikken SAOF, Januzzi JL, Granger CB, van Beurden A, et al. Zalunfiban at First Medical Contact for ST-Elevation Myocardial Infarction. NEJM Evid. 2026;5(1):EVIDoa2500268. First published online 10 November 2025.

Introduction

The interval between first medical contact and coronary reperfusion is an opportunity for treatment in ST-elevation myocardial infarction. Platelet-rich thrombus may continue to obstruct the infarct-related artery while a patient travels to the catheter laboratory. Oral antiplatelet treatment can be given early, but absorption and onset are not immediate, particularly during acute illness. A rapidly acting parenteral drug could bridge that interval.

Glycoprotein IIb/IIIa inhibitors block the final common pathway of platelet aggregation. Earlier intravenous agents can provide potent inhibition, but continuous administration, bleeding and thrombocytopenia complicate their use before hospital arrival. A single subcutaneous dose with rapid onset and short duration offers a more practical way to deliver this mechanism at first contact. Ease of administration, however, does not establish that the added platelet inhibition produces a favourable clinical balance.

Zalunfiban was developed for this purpose. CELEBRATE tested it alongside contemporary antithrombotic treatment in early-presenting patients intended for primary percutaneous coronary intervention. The trial combined clinically serious events with a biochemical infarct-size category in an ordered primary outcome. Understanding that scale is essential: a positive result does not necessarily mean fewer deaths or fewer major ischaemic events.¹

Synopsis

CELEBRATE was an international, phase III, double-blind, placebo-controlled trial recruiting in eight countries between April 2021 and May 2025. The executive committee and CeleCor Therapeutics, the sponsor, jointly designed and managed the study. An independent data and safety monitoring board oversaw safety, masked committees adjudicated clinical outcomes, and independent core laboratories assessed angiography and electrocardiograms. Two independent statisticians performed and validated the analyses.^{1,2}

Patients were eligible with persistent ischaemic chest pain lasting more than ten minutes, at least 2 mm of new ST elevation in two adjacent leads, and no more than four hours from symptom onset to the diagnostic electrocardiogram. They required planned transport to a participating primary PCI service and a reported weight of 52–130 kg. Men were eligible from age 18. The supplementary criteria restricted women to those who were postmenopausal or surgically sterile and at least 50 years old, or 55 at Czech sites.

Exclusions included out-of-hospital cardiac arrest requiring resuscitation, cardiogenic shock, dialysis, current oral anticoagulation or thrombolytic treatment, recent major surgery, trauma or bleeding, previous ischaemic or haemorrhagic stroke and known severe anaemia requiring regular transfusion. Consent procedures followed local law and could include witnessed verbal consent, a short written process or an approved exception, with subsequent written consent when possible.

Randomisation used permuted blocks stratified by country and allocated patients in a 1:1:1 ratio to a single subcutaneous injection of zalunfiban 0.11 mg/kg, zalunfiban 0.13 mg/kg or placebo. Identical sealed kits concealed allocation from patients, emergency personnel, investigators and the sponsor. Trained staff administered treatment at home, in an ambulance or at hospital. Aspirin, heparin and oral P2Y12 inhibitors could be given according to local practice. Revascularisation and subsequent antithrombotic treatment remained at the treating team's discretion.

The primary efficacy outcome was a seven-category ordinal scale based on the worst event experienced. From worst to best, the categories were death, stroke, recurrent myocardial infarction, acute stent thrombosis, new or recurrent heart failure, a troponin-defined larger infarct, and none of these outcomes. Most events were assessed through 30 days; acute stent thrombosis concerned the first 24 hours after PCI. The biochemical category used high-sensitivity troponin T at least thirty times the upper limit of normal in the specified early sampling window.

The two active doses were pooled for the primary comparison, as planned. A proportional-odds model estimated the odds of being in a worse outcome category, adjusted for country, age, baseline heart rate, systolic pressure and anterior versus non-anterior infarct location. Missing outcome components were addressed by multiple imputation. This was an ordinal regression analysis, not a pairwise win-ratio analysis.

Three key secondary efficacy outcomes followed a sequential testing hierarchy: corrected TIMI frame count at the initial angiogram, ST-segment resolution one hour after PCI, and bailout intravenous glycoprotein IIb/IIIa or P2Y12 inhibitor use. Prespecified exploratory outcomes included initial TIMI grade 2 or 3 flow and

ST-segment resolution before angiography. The primary safety endpoint was GUSTO severe or life-threatening bleeding through 30 days. Other safety outcomes included less severe bleeding, thrombocytopenia and injection-site reactions.

The planned sample was 2,334 patients for 90% power, increased to 2,499 to allow for attrition. Ultimately, 2,467 were randomised: 853 to the lower dose, 818 to the higher dose and 796 to placebo. The supplement attributes stopping slightly below the target to lower-than-expected attrition. All randomised patients entered the efficacy analysis with imputation where needed. Thirty-day clinical follow-up was completed in 1,637 of 1,671 active-treatment patients and 780 of 796 controls. Safety analyses used treatment received and included 1,664 and 795 patients, respectively.

Mean age was 63 years, approximately 21% were women and 95% were White. Around 73% were recruited in the Netherlands, with most of the remainder from other European countries. Approximately 87% were enrolled in an ambulance. Median time from symptoms to treatment was about 90 minutes, from study treatment to angiography 36 minutes, and from treatment to PCI 50 minutes. Approximately 95% received heparin plus an oral P2Y₁₂ inhibitor before or alongside study treatment. Radial access was used for approximately 95% of angiograms.

The primary ordinal outcome favoured pooled zalunfiban, with an adjusted odds ratio of 0.79, a 95% confidence interval of 0.65–0.98 and $P=0.028$. Lower odds indicate a more favourable distribution across the scale. The cumulative comparison through death, stroke, recurrent infarction, acute stent thrombosis and heart failure had an odds ratio of 0.87, with a 95% confidence interval of 0.65–1.16. The comparison that additionally included the troponin category was more favourable, with an odds ratio of 0.71 and a 95% confidence interval of 0.54–0.94.¹

Observed individual event rates, which differ from the imputed cumulative hierarchy, provide clinical context. Death occurred in 37 of 1,637 active-treatment patients, 2.3%, and seventeen of 780 controls, 2.2%. Recurrent infarction occurred in 1.9% versus 1.2%, and stroke in 0.7% versus 0.8%. Acute stent thrombosis occurred in three of 1,671 versus eight of 796 patients, 0.2% versus 1.0%. New or recurrent heart failure occurred in 6.5% versus 8.1%. These component comparisons were not independently established confirmatory treatment effects.

The troponin threshold was met in 1,357 of 1,588 evaluable active-treatment patients, 85.4%, and 668 of 755 controls, 88.5%. No component event was observed in 13.3% versus 9.8% of the relevant evaluable groups. Because the biochemical category was common, it contributed heavily to the distribution underlying the primary result.

Corrected TIMI frame count at initial angiography favoured zalunfiban, with medians of 109 versus 176 and a reported median difference of –66 frames, with a 95% confidence interval of –93 to –39; $P=0.012$. This passed the first key secondary test. One-hour post-PCI ST-segment resolution did not differ clearly, with $P=0.42$, so the confirmatory sequence stopped. Bailout drug use was 8.6% versus 9.5%, but was not formally tested after that gate closed. Exploratory initial TIMI grade 2 or 3 flow occurred in 52.0% versus 45.1%.

Severe or life-threatening bleeding occurred in twenty of 1,664 zalunfiban recipients, 1.2%, and six of 795 placebo recipients, 0.8%. The risk difference was 0.4 percentage points, with a 95% confidence interval of –0.4 to 1.2; $P=0.40$. GUSTO mild or moderate bleeding increased to 6.4% from 2.5%. Injection-site reactions occurred in 5.8% versus 0.6%, and platelet counts below 150,000/ μL in 8.1% versus 6.3%, without a clear difference in the latter comparison.

Critique

CELEBRATE establishes a statistically positive primary ordinal outcome and improved coronary flow before PCI. The masked randomised design, treatment at first contact and use alongside modern antithrombotic therapy are substantial strengths. The central interpretive issue is what the composite improvement means for patients, rather than whether the primary P value crossed its threshold.

The outcome places death at the worst end and a troponin-defined infarct at the least severe event end. This ordering acknowledges clinical priorities, but a proportional-odds model does not automatically make death dominate the statistical result. It estimates a common treatment effect across cumulative divisions of the scale. The large number of biochemical events can therefore strongly influence the overall estimate even when death and other serious outcomes show little difference.

An odds ratio of 0.79 should not be translated into 21% fewer deaths, 21% fewer major clinical events or a 21% reduction in each component. It describes the modelled shift across the ordered outcome. The individual cumulative odds ratios ranged from values close to one for death or stroke to a more favourable estimate when the troponin category was included. This pattern makes the common proportional-odds summary useful but incomplete. The supplied results do not provide enough detail to independently assess how well a single common effect describes every threshold.

The raw component outcomes reinforce that caution. Mortality was almost identical, recurrent infarction was numerically more frequent with zalunfiban, and the apparent

heart failure reduction was uncertain. Acute stent thrombosis was less frequent, but only eleven events occurred and the comparison was exploratory. A positive primary composite does not transform each favourable component into a separately proven benefit or make an unfavourable numerical difference disappear.

The troponin category is particularly important. More than 85% of evaluable patients in both groups crossed the threshold, so the label of a larger infarct applies to most participants. Troponin release is clinically informative, but a dichotomised laboratory value is not a direct measurement of scar size, ventricular function, symptoms or prognosis. Patients just above and far above the threshold enter the same category, while small differences around it can move patients between categories.

The threshold also changed during the trial. The supplement states that ten times the upper limit of normal was initially planned, but fewer patients than expected had values below that level. The executive committee increased the threshold to thirty times the upper limit to improve statistical power. This is a consequential change to the most frequent component of the primary outcome and should be visible when interpreting the result.²

The main report states that treatment allocation remained masked, which is reassuring, but the supplied results do not fully document the amendment's timing or the treatment-group information available to the decision-makers. The supplement asserts that exploratory peak troponin comparisons favoured zalunfiban irrespective of the threshold, without presenting a complete set of alternative primary analyses. That statement does not remove the need for transparent reporting of the amendment. The concern is interpretive and methodological; it is not evidence that the decision was improperly made.

The number needed to treat of approximately 29 quoted in the discussion comes from the 3.5-point difference in having none of the component outcomes among evaluable patients. It should not be presented as a number needed to prevent one death or one major clinical complication. Its meaning includes avoiding the troponin threshold and depends on the denominators and missing-data handling used for that descriptive comparison. The ordinal primary effect itself does not yield a single universal number needed to treat.

Clinical follow-up was good, with about 2% not completing the 30-day assessment, but biochemical and electrocardiographic data were missing more often. The troponin comparison included 1,588 active-treatment patients and 755 controls, while the post-PCI ECG analysis included 1,267 and 609. Multiple imputation preserves inclusion in

the primary analysis under assumptions about the missing values; it does not make those values observed. Given the biochemical component's influence, the robustness of those assumptions deserves attention before treating the effect size as definitive.

The difference between imputed hierarchical counts and observed event counts should also remain explicit. For example, the primary table includes 38 versus eighteen deaths, whereas the observed individual-event table gives 37 versus seventeen. These are different analysis presentations, not necessarily contradictory event adjudications. Mixing their denominators or summing the cumulative rows would produce an incorrect account of what happened.

The sequential secondary analysis is a strength because it limits false-positive claims. The initial flow measure improved, which supports the intended early pharmacological action. The next endpoint, ST-segment resolution after PCI, did not. Consequently, bailout drug use cannot be promoted to a confirmatory result. Exploratory flow grades and pre-PCI ECG findings are compatible with early benefit, but their intervals were not adjusted for multiple comparisons.

Timing may explain why an effect before angiography did not clearly persist in the post-PCI ECG endpoint: mechanical reperfusion can reduce differences created during the preceding interval. That is a plausible interpretation, not a demonstrated mechanism. Whether the early flow improvement translates into durable ventricular recovery or fewer later clinical events requires direct evidence beyond this short-term composite.

The safety conclusion also needs precision. Severe bleeding was not statistically increased, but the point estimate was higher and the confidence interval allowed an absolute increase of 1.2 percentage points. The trial did not establish equivalence or rule out an important rare bleeding cost. A statement that there was no demonstrated increase is more accurate than claiming the drug adds no severe bleeding risk.

Less severe bleeding clearly increased, by approximately four percentage points, and the broader BARC bleeding comparison also favoured placebo. These events matter to patients and can influence subsequent antithrombotic treatment, investigation and length of care. They should not be dismissed because they occupy a less severe category than intracranial or life-threatening bleeding. Injection-site reactions add another practical burden, although generally distinct from the major ischaemic outcomes the drug aims to prevent.

Pooling the two doses was planned and improves power for the drug-versus-placebo question. It does not identify the optimal dose, establish equal efficacy between doses or show that their bleeding profiles are interchangeable. A pooled benefit cannot automatically be assigned to either dose separately. Dose selection therefore requires evidence beyond the principal comparison reported here.

The prehospital setting is a major strength but also a source of selection. Patients had to be recognised early, meet relatively specific ECG and weight criteria, enter a participating pathway and satisfy the local consent process. Comprehensive screening data were unavailable, so the proportion of ordinary STEMI patients represented cannot be determined. The trial tests use by trained emergency personnel within an organised PCI system, not unrestricted administration to anyone with possible cardiac chest pain.

Generalisation to the sickest critical care patients is particularly limited. Cardiogenic shock and resuscitated cardiac arrest were exclusions, as were previous stroke, chronic anticoagulation and dialysis. Women were underrepresented and subject to additional age and reproductive criteria. Nearly all participants were White and most were treated in the Netherlands. These features do not undermine randomisation but narrow the population in whom the balance of benefit and bleeding was measured.

The treatment interval was also short and access was predominantly radial. A longer transfer could offer more time for early platelet inhibition to help, but it could also change concomitant treatment, diagnosis and exposure to bleeding risk. Greater benefit with longer delay is a hypothesis, not a finding from this comparison. Likewise, bleeding outcomes may differ where femoral access is common. The study did not randomise home or ambulance administration against administration on hospital arrival.

The sponsor's involvement was extensive, with independent oversight, masked adjudication and academic data access providing safeguards. The accompanying editorial appropriately treats the result as promising while asking for confirmation in hard clinical outcomes. The most defensible conclusion is that rapid subcutaneous platelet inhibition improved early reperfusion and the ordered composite, at a clear cost in non-severe bleeding; the magnitude of net patient-centred benefit remains uncertain.³

Body of Evidence

ATLANTIC randomised 1,862 patients with STEMI to ticagrelor in the ambulance or in the catheter laboratory. The median treatment separation was 31 minutes. Earlier administration did not improve the coprimary pre-PCI reperfusion endpoints, although stent thrombosis was less frequent in secondary analyses. This illustrates the difficulty of obtaining a rapid pre-PCI effect from an oral agent within a short transfer interval. It does not establish that adding another antiplatelet drug will improve net clinical outcomes.⁴

On-TIME 2 tested prehospital high-bolus tirofiban against placebo in 984 patients receiving aspirin, heparin and clopidogrel. Residual ST-segment deviation one hour after PCI was lower with tirofiban, while major bleeding was not clearly different. The trial supports the biological plausibility of early potent platelet inhibition, but it used an intravenous agent and a different background antiplatelet regimen. CELEBRATE tests a more convenient route in a later treatment era.⁵

FINESSE is a caution against assuming that earlier pharmacological reperfusion must improve clinical outcomes. It randomised 2,452 patients among facilitated strategies using early abciximab, with or without half-dose reteplase, and a primary PCI strategy with abciximab at the procedure. The 90-day clinical composite did not clearly differ. The comparator was not absence of platelet inhibition: all groups received abciximab around PCI. The study demonstrates that timing and physiological improvement require their own outcome evidence.⁶

CHAMPION PHOENIX provides evidence for another rapidly acting parenteral antiplatelet strategy. Among 11,145 patients undergoing urgent or elective PCI, intravenous cangrelor reduced a 48-hour ischaemic composite from 5.9% to 4.7% compared with clopidogrel, without a demonstrated increase in severe bleeding. This supports the potential value of prompt platelet inhibition, but the drug, comparator, clinical setting and timing differ substantially from prehospital zalunfiban.⁷

The small CEL-02 study established pharmacological feasibility for subcutaneous RUC-4, subsequently named zalunfiban. Twenty-seven STEMI patients received escalating doses, with rapid, dose-related platelet inhibition by fifteen minutes. Its size and open-label design could not establish clinical efficacy or reliably characterise uncommon harm. CELEBRATE advances the programme to a randomised phase III outcome comparison, while leaving the clinical importance of its composite shift and the best dose incompletely resolved.⁸

Taken together, the evidence supports a plausible and practical way to intensify treatment before PCI. It also shows why a favourable early flow measure is insufficient on its own. Confirmation of durable benefit in major clinical outcomes, alongside a clear account of bleeding, is the next important question.

Should zalunfiban routinely be added at first medical contact for STEMI?

CELEBRATE is promising but does not yet establish a clear net benefit in hard clinical outcomes. Its positive ordinal composite and early flow improvement must be weighed against increased mild-to-moderate bleeding and uncertainty about the contribution of the troponin endpoint.

References

1. Van't Hof AWJ, Gibson CM, Rikken SAOF, Januzzi JL, Granger CB, van Beurden A, et al. Zalunfiban at First Medical Contact for ST-Elevation Myocardial Infarction. *NEJM Evid.* 2026;5(1):EVIDoa2500268.
2. van't Hof AWJ, et al. Supplementary Appendix to Zalunfiban at First Medical Contact for ST-Elevation Myocardial Infarction. *NEJM Evid.* 2026.
3. d'Entremont MA, Jolly SS. Subcutaneous Glycoprotein IIb/IIIa Inhibitor for ST-Elevation Myocardial Infarction - Teaching an Old Drug New Tricks. *NEJM Evid.* 2026;5(1):EVIDe2500280.
4. Montalescot G, van 't Hof AW, Lapostolle F, Silvain J, Lassen JF, Bolognese L, et al. Prehospital ticagrelor in ST-segment elevation myocardial infarction. *N Engl J Med.* 2014;371(11):1016-1027.
5. Van't Hof AW, Ten Berg J, Heestermans T, Dill T, Funck RC, van Werkum W, et al. Prehospital initiation of tirofiban in patients with ST-elevation myocardial infarction undergoing primary angioplasty (On-TIME 2): a multicentre, double-blind, randomised controlled trial. *Lancet.* 2008;372(9638):537-546.
6. Ellis SG, Tendera M, de Belder MA, van Boven AJ, Widimsky P, Janssens L, et al. Facilitated PCI in patients with ST-elevation myocardial infarction. *N Engl J Med.* 2008;358(21):2205-2217.
7. Bhatt DL, Stone GW, Mahaffey KW, Gibson CM, Steg PG, Hamm CW, et al. Effect of platelet inhibition with cangrelor during PCI on ischemic events. *N Engl J Med.* 2013;368(14):1303-1313.
8. Bor WL, Zheng KL, Tavenier AH, Gibson CM, Granger CB, Bentur O, et al. Pharmacokinetics, pharmacodynamics, and tolerability of subcutaneous administration of a novel glycoprotein IIb/IIIa inhibitor, RUC-4, in patients with ST-segment elevation myocardial infarction. *EuroIntervention.* 2021;17(5):e401-e410.

IMPROVE-multi

**Saugel B, Meidert AS, Brunkhorst FM, Bischoff R, Esser J, Mattis M, et al.
Individualized Perioperative Blood Pressure Management in Patients Undergoing
Major Abdominal Surgery: The IMPROVE-multi Randomized Clinical Trial. JAMA.
2025;334(21):1893-1904.**

Introduction

Blood pressure is measured throughout anaesthesia because it provides an immediate warning of circulatory disturbance. Low pressure can accompany vasodilatation, blood loss, impaired cardiac function or excessive anaesthetic effect, and observational studies associate greater hypotension exposure with postoperative organ injury. The unresolved question is how much of that association can be prevented by targeting a higher pressure, particularly when doing so requires more vasopressor treatment.¹

An individualised target seems attractive because patients arrive for surgery with different usual blood pressures. A mean arterial pressure of 65 mmHg might represent a modest reduction for one person and a large reduction for another. Yet a preoperative pressure is not necessarily a measure of the pressure required for adequate organ perfusion during anaesthesia. It may reflect anxiety, medication, sleep disruption, chronic vascular disease or normal circadian variation.

IMPROVE-multi used repeated overnight measurements to define a patient's reference mean arterial pressure. This was intended to provide a more stable baseline than a single reading shortly before induction. The hypothesis was that maintaining perioperative pressure at least at the patient's mean nighttime level would reduce kidney injury, myocardial injury, cardiac arrest or death compared with a routine target of at least 65 mmHg.¹

The trial is important because it tested a reproducible version of personalised pressure management, rather than merely comparing two fixed thresholds. It also helps distinguish the value of avoiding profound hypotension from the proposed benefit of raising pressure substantially above an already well-maintained conventional minimum. This is the blood-pressure IMPROVE-multi trial, distinct from the earlier IMPROVE ventilation study.

Synopsis

IMPROVE-multi was an investigator-initiated superiority trial at fifteen German university hospitals. Recruitment occurred from February 2023 to April 2024, with final follow-up in July 2024. The German Research Foundation funded the study and was reported to have no role in its conduct, analysis or publication decision after the funding review. All participants gave written informed consent, and an independent monitoring board reviewed safety and the planned interim analysis. The study was described as single-blind, although treating anaesthesiologists necessarily knew the assigned target.¹

Eligible patients were at least 45 years old, were scheduled for elective major abdominal surgery under general anaesthesia expected to last at least ninety minutes, and had at least one additional risk criterion. These criteria included older age, cardiovascular or respiratory disease, diabetes, obesity, smoking, renal impairment, immunodeficiency or an expected anaesthetic duration of at least three hours. The definitions were broad: renal impairment could include estimated filtration below 90 mL/min/1.73 m², and solid-organ cancer could fulfil the immunodeficiency criterion. Emergency surgery, sepsis, dialysis, specified transplant procedures or history, and inability to obtain overnight measurements were among the exclusions.^{1,2}

Before randomisation, all participants underwent automated blood-pressure measurement every thirty minutes for one night. After excluding artefacts, measurements between midnight and 06:00 were averaged. Monitoring could occur at home or in hospital without a fixed required interval before surgery. In practice, approximately 95% were monitored in hospital, usually the preceding night, and the median number of usable nighttime readings was twelve.

Patients were randomised centrally in equal proportions, using variable blocks stratified by centre, before induction. In the individualised group, the target was a mean arterial pressure at least as high as the preoperative nighttime mean. Targets were bounded at 65 mmHg and 110 mmHg: lower nighttime means still led to a minimum target of 65, and means above 110 led to a target of at least 110. Control patients had a target of at least 65 mmHg. The intervention lasted from the start of induction until two hours after surgery.¹

The trial prescribed pressure targets without a uniform treatment algorithm. Anaesthesiologists selected fluids, vasopressors and other interventions according to clinical judgement. No systematic advanced haemodynamic monitoring was mandated. Blood pressure was measured invasively or with an oscillometric cuff according to local practice. The analysis quantified hypotension by integrating its depth and duration below specified thresholds, rather than counting only whether a low reading occurred.

The primary endpoint was any acute kidney injury, acute myocardial injury, nonfatal cardiac arrest or death within seven postoperative days. Kidney injury required at least a 50% creatinine increase from baseline or kidney replacement therapy. Myocardial injury used high-sensitivity troponin criteria or myocardial infarction. The supplementary definition required a postoperative troponin above the assay's 99th percentile and a specified increase from baseline. This biomarker endpoint was broader than clinically diagnosed myocardial infarction.^{1,2}

Twenty-two secondary outcomes covered the composite and its components at other time points, infection, readmission, discharge and a thirty- and ninety-day composite of kidney replacement therapy, myocardial infarction, cardiac arrest or death. The early and later composites therefore differed: later outcomes used myocardial infarction rather than the broader acute myocardial injury definition.

The trial assumed a primary-event rate of 25% with routine care and 17% with individualised care. It sought approximately 1,144 evaluable patients for 90% power, with 1,272 enrolled to allow for losses. A planned interim analysis used an O'Brien–Fleming spending function; the trial continued after that analysis. The final two-sided significance threshold was 0.049. The primary comparison used a continuity-corrected test of proportions, with centre-adjusted mixed-effects and per-protocol sensitivity analyses. Complete-case analysis was used without statistical imputation.^{1,2}

Of 7,621 patients screened, 1,272 were enrolled and 1,142 randomised, 571 per group. Most prerandomisation exclusions arose from inadequate overnight measurements or technical problems. Eight patients were excluded after randomisation, leaving 567 in each group for the primary analysis. Median age was approximately 66 years, 34% were women and three-quarters had solid-organ cancer. Around half had a history of hypertension. Although approximately two-thirds met the broad renal-risk criterion, only about 10% had estimated filtration below 60 mL/min/1.73 m². General and urological procedures predominated.¹

Median nighttime mean arterial pressure was 84 mmHg in the individualised group and 82 mmHg in controls. Approximately three-quarters had arterial-catheter monitoring. Individualised management achieved a higher pressure distribution and less exposure below 65 mmHg: median area below that threshold was 6 versus 48 mmHg·minutes. Median time at or above the patient's nighttime mean was 79% versus 57%. Median crystalloid administration was 3 L in both groups. Norepinephrine was used in 94.5% versus 87.3%, with median cumulative doses among recipients of 1.53 versus 1.25 mg.¹

The primary endpoint occurred in 190/567 patients receiving individualised care, 33.5%, and 173/567 receiving routine care, 30.5%. The relative risk was 1.10, 95% confidence interval 0.93–1.30; $P=0.31$. The absolute difference was 3.00 percentage points, 95% confidence interval –2.61 to 8.60. Centre-adjusted and per-protocol analyses gave similar conclusions. The per-protocol analysis excluded eight intervention patients with major deviations, principally incorrect calculation of their nighttime target.¹

Acute kidney injury occurred in 72/566 versus 58/567 patients, 12.7% versus 10.2%, and myocardial injury in 153/558 versus 138/561, 27.4% versus 24.6%. Neither difference was statistically significant. There were two versus no nonfatal arrests and no versus four deaths within seven days. At ninety days, the later composite occurred in 32/566 versus 20/567, 5.7% versus 3.5%, while mortality was sixteen patients in each group. None of the reported secondary comparisons established benefit. Subgroup analyses based on degrees of preoperative hypertension did not identify a convincing interaction.^{1,2}

Critique

The study tested a specific, implementable target-selection method and achieved meaningful treatment separation. Overnight measurement reduced dependence on a single potentially anxious preoperative reading. Central randomisation, centre stratification, a prespecified primary endpoint, a controlled interim analysis and near-complete follow-up strengthen the comparison. The negative result cannot readily be dismissed as a trial in which the two groups received indistinguishable pressure management.

The individualised strategy reduced hypotension exposure and increased vasopressor use without improving the primary outcome. This is a direct challenge to the assumption that returning a patient to an overnight reference pressure necessarily protects organs during anaesthesia. Mean arterial pressure is an important circulatory variable, but it is not a complete measure of blood flow, oxygen delivery or microcirculatory function. Raising pressure may correct vasodilatation while leaving hypovolaemia, myocardial dysfunction, anaemia or other causes of injury unresolved.^{1,3}

A central limitation lies in the target's biological meaning. Normal sleep and general anaesthesia differ in metabolic demand, vascular tone, ventilation and medication exposure. A mean nighttime pressure has not been validated as the lower boundary of renal or coronary autoregulation during surgery. Repeated measurement improves the precision of the reference number without proving that the number is the correct treatment target. The trial rejects this particular inference more directly than it rejects the broad idea of individualised haemodynamic care.

The measurement setting further complicates the concept of a normal baseline. Almost all overnight recordings occurred in hospital, usually shortly before surgery. Sleep disturbance, anxiety, discomfort and cuff inflation may have affected the values. Home monitoring might better reflect ordinary physiology, but would require additional organisation and would still not validate the link between baseline pressure and intraoperative perfusion requirements. The logistical losses before randomisation also demonstrate that overnight profiling carries a practical burden.

Control care was an active and relatively successful hypotension-avoidance strategy. Profound or prolonged hypotension was uncommon, even among controls. The median area below 65 mmHg was low in both groups, although lower with individualised treatment. The study therefore evaluates additional pressure elevation beyond a well-maintained conventional threshold. It does not identify the pressure below which injury becomes likely, test deliberate prolonged hypotension, or justify withholding treatment for shock. Interpreting it as evidence that blood pressure does not matter would exceed the comparison.

The confidence interval is clinically informative. It permits a modest absolute benefit of about 2.6 percentage points, but excludes the anticipated eight-point reduction and includes appreciable harm. The control event rate was actually higher than expected, so this is not primarily a trial undermined by unexpectedly few composite events. However, death and cardiac arrest were rare. The study cannot establish equivalent effects on those outcomes merely because their comparisons were nonsignificant.

The composite was dominated by biomarker-defined myocardial injury and creatinine-defined kidney injury. These outcomes are relevant, but their frequency and clinical meaning differ from arrest or death. The acute kidney injury definition did not include isolated oliguria or the smaller absolute creatinine increase used in some other classifications. The myocardial injury definition was not synonymous with infarction. Readers comparing event rates across trials must therefore consider the definitions rather than treating all postoperative organ-injury composites as interchangeable.²

Missing-data handling also qualifies the claim of complete primary ascertainment. The supplement states that missing postoperative creatinine or troponin measurements were treated as not elevated, while absent baseline values could make an individual component unavailable. Thus, all patients could have a composite classification despite some missing component data and unmeasured postoperative values. This is different from complete serial biochemical surveillance. If missing measurements were related to illness or discharge, the assumption could affect sensitivity for injury; the available report does not establish its direction.²

Treatment was deliberately left to clinical judgement. This enhances relevance to anaesthetic practice, but the intervention becomes a policy of pursuing a higher target through whatever measures clinicians choose. It cannot isolate the causal effect of pressure from the effects of the fluids or drugs used to achieve it. More norepinephrine may be a cost of the strategy, but the trial does not prove that norepinephrine caused the numerically greater kidney or myocardial injury. Systematic cardiac-output measurements would have helped describe mechanisms, without necessarily changing the clinical answer.

The modified intention-to-treat population excluded only eight randomised patients, evenly distributed, which is unlikely to dominate the result but should still be stated. The per-protocol comparison was also neutral. Analyses restricted by the proportion of time above target are less secure because achieved pressure is a postrandomisation variable influenced by patient stability and treatment response. Moreover, the supplementary table appears to reverse the stated risk-ratio direction relative to its displayed event proportions. Those estimates should not be used to manufacture an adherence-related benefit.

Hypertension is an obvious candidate modifier, but the subgroup findings do not support routinely raising targets for every patient with higher preoperative pressure. At the same time, the most extreme hypertension categories were small and their intervals wide. Absence of a significant interaction is not proof of a uniform effect across every vascular phenotype. Chronic antihypertensive continuation or withdrawal was not recorded, further limiting interpretation of how the preoperative reference related to medication exposure.

The practical conclusion is narrower and stronger than a general rejection of personalised care. In selected elective abdominal surgical patients managed in university hospitals, an overnight-pressure target added monitoring and vasopressor exposure without demonstrated organ protection. The findings favour retaining attention to the cause of hypotension and the patient's broader physiology rather than assuming that a higher personalised pressure number is itself a validated therapeutic objective.

Body of Evidence

INPRESS provided an influential positive precedent. It randomised 298 high-risk patients, with 292 included in its modified intention-to-treat analysis, to systolic pressure within 10% of a resting reference or standard treatment triggered by systolic pressure below 80 mmHg or a reduction exceeding 40%. The seven-day composite of systemic inflammation and organ dysfunction occurred in 38.1% versus 51.7%, relative risk 0.73, 95% confidence interval 0.56–0.94. However, its target, comparator, intervention duration and endpoint differed from IMPROVE-multi. The latter's conventional target of at least 65 mmHg cannot be assumed equivalent to the earlier systolic-trigger strategy.⁴

The BBB trial subsequently tested fixed mean-pressure targets of at least 75 versus 60 mmHg in patients at cardiovascular risk. Among 451 analysed patients from 458 randomised, the higher target shortened exposure below 65 mmHg but did not significantly reduce its composite of myocardial injury, major cardiovascular events or acute kidney injury. Event rates were 48% and 52%, with a risk difference of –4.2 percentage points and a confidence interval spanning –13 to 5. This anticipated the separation between better pressure control and uncertain clinical benefit seen in IMPROVE-multi.⁵

POISE-3 provided a much larger comparison in 7,490 patients receiving chronic antihypertensive treatment. Its hypotension-avoidance strategy combined an intraoperative target of at least 80 mmHg with selective withholding of antihypertensive drugs; the comparison used a target of at least 60 mmHg with continued medication. The thirty-day vascular composite occurred in 13.9% and 14.0%, hazard ratio 0.99, 95% confidence interval 0.88–1.12. Because medication management also differed, it was a trial of perioperative strategies rather than pressure alone. Nevertheless, it gave no general support to pursuing the higher target.⁶

BP-CARES, published in 2025, studied high-risk abdominal surgery at three Chinese sites. Among 1,477 analysed patients, targeting at least 80 mmHg reduced hypotension exposure but did not reduce cardiovascular events compared with a conventional target based on at least 65 mmHg or 60% of baseline, whichever was higher. The composite occurred in 14.5% versus 13.6%, relative risk 1.07, 95% confidence interval 0.83–1.38. Its concordance with IMPROVE-multi extends the finding across healthcare settings and different ways of selecting the comparison target.⁷

PRETREAT examined a different personalised approach at two Dutch tertiary hospitals: mean-pressure targets of at least 70, 80 or 90 mmHg according to predicted hypotension risk. Recruitment stopped for futility after 3,247 participants. Six-month disability scores were similar, with a mean difference of –0.5 points and a 95% credible interval of –1.9 to

0.9, well short of the prespecified five-point clinically important difference. Unlike IMPROVE-multi, this trial directly addressed longer-term function, making it difficult to argue that neutral organ-injury findings necessarily conceal a substantial functional benefit from routine higher targets.⁸

Stop-or-Not adds a complementary perspective. In 2,222 patients taking renin–angiotensin system inhibitors, continuing treatment increased intraoperative hypotension compared with preoperative discontinuation, 54% versus 41%, but the twenty-eight-day composite of mortality and major complications was 22% in both groups. This does not prove hypotension harmless: medication continuation changes more than blood pressure, and observed episodes were managed clinically. It does show why reduction of a physiological event cannot automatically be accepted as proof of improved patient outcomes.⁹

Together, these trials weaken the premise that routinely raising perioperative pressure above an adequately maintained conventional minimum produces organ protection. IMPROVE-multi specifically adds that overnight measurement and personalised targets do not rescue that premise. The remaining challenge is to identify when low pressure reflects a treatable perfusion deficit and which intervention addresses its cause, while avoiding unsupported escalation simply to reproduce a preoperative number.

Should overnight blood-pressure measurements routinely determine higher perioperative MAP targets for high-risk abdominal surgery?

No. IMPROVE-multi reduced hypotension exposure but did not improve postoperative outcomes compared with targeting MAP of at least 65 mmHg; it does not establish that profound or prolonged hypotension is safe.

References

1. Saugel B, Meidert AS, Brunkhorst FM, Bischoff R, Esser J, Mattis M, et al. Individualized Perioperative Blood Pressure Management in Patients Undergoing Major Abdominal Surgery: The IMPROVE-multi Randomized Clinical Trial. JAMA. 2025;334(21):1893-1904.
2. Saugel B, Meidert AS, Brunkhorst FM, Bischoff R, Esser J, Mattis M, et al. Supplementary methods and results to Individualized Perioperative Blood Pressure Management in Patients Undergoing Major Abdominal Surgery: The IMPROVE-multi Randomized Clinical Trial. JAMA. 2025;334(21):1893-1904.

3. Legrand M, Lamontagne F, Pirracchio R. Perioperative Outcomes-The Limits of Blood Pressure-Centered Strategies. *JAMA*. 2025;334(21):1885-1887.
4. Futier E, Lefrant JY, Guinot PG, Godet T, Lorne E, Cuvillon P, et al. Effect of Individualized vs Standard Blood Pressure Management Strategies on Postoperative Organ Dysfunction Among High-Risk Patients Undergoing Major Surgery: A Randomized Clinical Trial. *JAMA*. 2017;318(14):1346-1357.
5. Wanner PM, Wulff DU, Djurdjevic M, Korte W, Schnider TW, Filipovic M. Targeting Higher Intraoperative Blood Pressures Does Not Reduce Adverse Cardiovascular Events Following Noncardiac Surgery. *J Am Coll Cardiol*. 2021;78(18):1753-1764.
6. Marcucci M, Painter TW, Conen D, Lomivorotov V, Sessler DI, Chan MTV, et al. Hypotension-Avoidance Versus Hypertension-Avoidance Strategies in Noncardiac Surgery : An International Randomized Controlled Trial. *Ann Intern Med*. 2023;176(5):605-614.
7. Zhao B, Zhang J, Xie Y, Wu Z, Guo G, Lei S, et al. Intensive vs Conventional Intraoperative Blood Pressure Management on Cardiovascular Events After Major Abdominal Surgery: The BP-CARES Randomized Trial. *J Am Coll Cardiol*. 2025;86(12):892-906.
8. Kant M, van Klei WA, Hollmann MW, de Klerk ES, Otterspoor LC, Besselink MG, et al. Proactive vs Reactive Treatment of Hypotension During Surgery: The PRETREAT Randomized Clinical Trial. *JAMA*. 2025;334(21):1905-1914.
9. Legrand M, Falcone J, Cholley B, Charbonneau H, Delaporte A, Lemoine A, et al. Continuation vs Discontinuation of Renin-Angiotensin System Inhibitors Before Major Noncardiac Surgery: The Stop-or-Not Randomized Clinical Trial. *JAMA*. 2024;332(12):970-978.

PRINCE

Greco M, Lombardi G, Brusasco C, Pieri M, Roasio A, Monaco F, et al. Effect of Remote Ischemic Preconditioning on Myocardial Injury in Noncardiac Surgery: The PRINCE Randomized Clinical Trial. *Circulation*. 2025;152(17):1194-1205.

Introduction

Remote ischaemic preconditioning is an appealingly simple intervention: briefly interrupt blood flow to a limb, then restore it, in the hope that protective signals reduce injury elsewhere. Repeated cuff inflation is inexpensive and can be performed during preparation for surgery. Its biological rationale is that transient, controlled ischaemia activates neural and circulating pathways that increase tolerance to subsequent tissue stress. The clinical question is whether that response is sufficiently reliable to protect patients undergoing major surgery.

The literature has repeatedly moved from promising small studies to less encouraging large trials. In cardiac surgery, ERICCA and RIPHeart did not show improvement in major clinical outcomes. Their findings raised questions about whether anaesthetic drugs, particularly propofol, interfered with the conditioning response. Such a possibility is biologically interesting, but it also creates a demanding test for clinical translation: an intervention must work within a reproducible treatment setting, not only under an assumed set of ideal conditions.^{1,2}

Noncardiac surgery remained a plausible setting for benefit. The PIXIE trial in emergency hip-fracture surgery found less postoperative myocardial injury with preconditioning, although it was not large enough to establish an overall improvement in major cardiovascular events or survival. Differences in surgery, timing, masking and biomarker definitions left uncertainty about whether its result would generalise.³

PRINCE tested remote ischaemic preconditioning in a larger, international, sham-controlled population undergoing intermediate- or high-risk noncardiac surgery. It deliberately avoided propofol for induction and maintenance, addressing one proposed explanation for earlier neutral trials. The primary question was whether three cycles of limb ischaemia after anaesthetic induction would reduce postoperative myocardial injury.⁴

Synopsis

PRINCE was a pragmatic, parallel-group, double-blind trial conducted at 25 hospitals in eight countries. Recruitment extended from April 2015 to December 2024. The Italian Ministry of Health funded the study without a role in design, conduct, analysis or publication. Patients gave written informed consent. An independent monitoring committee performed a blinded interim review after approximately half the planned recruitment.⁴

Eligible patients were older than 50 years, were undergoing intermediate- or high-risk noncardiac surgery under general anaesthesia, and were receiving antiplatelet treatment before surgery as a marker of cardiovascular risk. Important exclusions included myocardial infarction in the preceding month, unstable or ongoing angina, pregnancy and peripheral vascular disease affecting the limbs. The population was therefore selected for cardiovascular risk but excluded some patients with particularly unstable disease.

Randomisation was computer generated within the electronic case-report system. A designated researcher who was not part of routine operating-room care received the allocation and performed the procedure. Patients, the clinical team, outcome assessors and the trial statistician were masked. The cuff and pressure display were covered to support masking, while a sham procedure matched the sequence and duration of treatment.⁵

The cuff was placed on an upper or lower limb before anaesthetic induction. Immediately after induction, the intervention consisted of three cycles of inflation to 200 mmHg for five minutes, followed by five minutes of deflation. The supplement states that this exceeded systolic pressure by at least 50 mmHg in all treated patients. The sham cuff's release valve remained open while inflation was mimicked. The whole sequence was performed before surgery began.

Propofol was to be avoided for both induction and maintenance. Other anaesthetic and perioperative management remained with the treating team. This allowed the trial to examine conditioning in a largely propofol-free environment without specifying one alternative anaesthetic regimen. The intervention was an addition to usual surgical and anaesthetic care, not a replacement for cardiovascular risk management.

The primary endpoint was postoperative myocardial injury, defined by at least one cardiac troponin value above the local assay's 99th-percentile upper reference limit during the first three postoperative days. At least one postoperative measurement was requested, with daily sampling for three days where possible. Preoperative troponin was

not routinely measured. The endpoint therefore identified an elevated postoperative concentration without requiring a demonstrated rise from an individual preoperative baseline.

Secondary outcomes included myocardial infarction and stroke within 30 days, acute kidney injury within seven days, ICU admission, hospital length of stay and 30-day all-cause mortality. Myocardial infarction required clinical evidence consistent with ischaemia in addition to the appropriate biomarker or other diagnostic criteria. The investigators also examined troponin elevations above five times the upper reference limit, major postoperative complications and possible treatment-related adverse events. Masked telephone follow-up assessed outcomes at 30 days, with additional attempts through general practitioners or correspondence when needed.^{4,5}

The sample-size calculation assumed a 22% myocardial-injury rate in controls and sought a seven percentage point absolute reduction, with 80% power and a two-sided 5% significance level. The calculated requirement was 546 patients per group, rounded to 1,100 overall. The primary analysis compared available outcome proportions without imputation. Per-protocol, as-treated, adjusted logistic and missing-data sensitivity analyses were additional assessments. Secondary intervals were not adjusted for multiplicity.

The flow diagram records 2,318 patients assessed and 1,217 randomised, 602 to conditioning and 615 to sham. Four later withdrew permission to use their data, leaving 599 and 614 in the retained population. This distinction is compressed in the abstract's description of 1,213 randomised patients. Primary troponin outcomes were available for 566 intervention patients and 596 controls.

Mean age was 70 years and approximately 40% were women. About 58% had ASA physical status of at least III, approximately 80% had treated hypertension and one-fifth had a previous myocardial infarction. Nearly all were taking antiplatelet therapy, predominantly aspirin. Around 44% underwent high-risk procedures; abdominal or intrathoracic operations accounted for one-third. The population was mainly White and was largely recruited from European settings.

Three cycles were completed in 568/588 intervention patients and 608/612 controls with procedure-completion data. The upper limb was used in 502/584 and 512/612, respectively; the remainder had lower-limb placement. Median time from induction to incision was 30 minutes in both groups. Propofol was administered to a small minority despite the intended restriction. Crossovers and accidental unmasking were uncommon.

At least one postoperative troponin was obtained in 94.5% of intervention patients and 97.1% of controls. Only 251/599 and 262/614 had measurements on all three days. One postoperative value was available in 202 and 226 patients. Later missing samples were attributed to logistical reasons, weekends, discharge and other circumstances; missingness was not explained solely by early recovery and discharge.⁵

Myocardial injury occurred in 215/566 patients (38.0%) with conditioning and 223/596 (37.4%) with sham: relative risk 1.02, 95% confidence interval 0.88–1.18; $P=0.84$. Results were similar in the principal sensitivity analyses. Troponin above five times the upper reference limit occurred in 24 versus 27 patients, relative risk 0.94, 95% confidence interval 0.55–1.60.

Myocardial infarction occurred in 9/595 versus 16/610 patients, relative risk 0.58, 95% confidence interval 0.26–1.29. Acute kidney injury occurred in 17/597 versus 16/613. Median hospital stay was five days in both groups, and ICU admission occurred in 171/593 versus 162/613. Thirty-day deaths were 10/599 versus 8/611, relative risk 1.28, 95% confidence interval 0.51–3.21. None of these comparisons established benefit.

Limb petechiae occurred in 10/596 intervention patients and 1/608 controls. Unplanned hospital readmission occurred in 34/562 versus 20/576, relative risk 1.74, 95% confidence interval 1.02–2.99. Other adverse-event comparisons were not clearly different. A subgroup analysis showed more myocardial injury among patients receiving lower-limb conditioning, but did not establish a clinically actionable explanation for the overall result.

Critique

PRINCE materially weakens the case for routinely applying this conditioning regimen during noncardiac surgery. The primary estimate was close to no effect, and its interval does not support the large relative benefit anticipated when planning the trial. The absence of a positive result is especially informative because the design used sham treatment, masked clinical care and largely avoided propofol. Several common explanations for earlier negative studies were therefore addressed directly.⁴

The primary endpoint nevertheless has a substantial measurement limitation. Without a preoperative troponin value, an elevated postoperative concentration cannot always be identified as new injury caused by the perioperative period. Some patients may have chronic elevation. Randomisation should balance that tendency on average, but it cannot make the endpoint specific to acute ischaemic injury. Including elevations that the

intervention could not plausibly prevent may dilute an effect on the intended mechanism.

The endpoint also does not distinguish all relevant causes of troponin release. Myocardial injury can accompany perioperative supply-demand imbalance, sepsis, arrhythmia and other systemic illness. The investigators performed a sensitivity analysis excluding several alternative causes, with a similarly neutral result. That supports the main finding, but it does not replace preoperative and serial postoperative measurements or formal adjudication of the mechanism of injury. A prognostically important biomarker remains an imperfect surrogate for treatment-responsive ischaemia.⁵

Sampling completeness is a separate issue from whether any measurement was available. More than 95% of the retained population had at least one postoperative value, but only about two-fifths had all three requested daily measurements. A single normal result may miss a later rise, while a single elevated result cannot describe its trajectory. Similar sampling patterns between groups reduce concern about gross differential surveillance; they do not guarantee equal sensitivity or establish that missing values were innocuous. The absence of all troponin measurements was also somewhat more frequent with conditioning.

The missing-data sensitivity analyses tested useful but limited scenarios. Assigning all missing intervention outcomes to injury and all missing control outcomes to no injury produced a result less favourable to conditioning. Assigning all missing values in both groups to injury also retained a neutral estimate. These analyses show that specified adverse assumptions do not reveal a hidden benefit. They do not recover unmeasured biomarker trajectories or fully address the possibility that the endpoint detected different types of myocardial injury across hospitals.

Assay variation further affects the meaning of the event rate. Centres used different troponin assays and reference limits, including conventional and high-sensitivity methods. Local thresholds are pragmatic and allow broad participation, but the same nominal endpoint need not represent identical injury severity everywhere. The observed myocardial-injury rate of approximately 38% was higher than the 22% assumed for planning. That increases the number of observed events, while also raising questions about whether the detected abnormalities match the clinical construct used in the initial effect-size assumption.

The contrast between frequent biomarker elevation and infrequent clinical events is important. Only 25 myocardial infarctions and 18 deaths were observed. The numerical

reduction in infarction is compatible with benefit, but its interval also includes no benefit and potential harm. It cannot be promoted to a positive clinical conclusion after a neutral primary result. Conversely, the trial was not large enough to exclude a modest effect on rare clinical outcomes. The appropriate conclusion is no demonstrated improvement, not proof that every possible cardiovascular effect is exactly zero.

Intervention fidelity was generally good. Most patients completed the three-cycle sequence, the time window was operationally practical and contamination was limited. The sham procedure is a meaningful strength because enthusiasm for a low-cost, technically simple intervention could otherwise alter perioperative behaviour or outcome assessment. Petechiae might reveal allocation in some patients, but their small number is unlikely to account for the overall near-identical biomarker rates.

Avoiding propofol strengthens the test of one prominent mechanistic hypothesis. It also narrows direct applicability to routine services that commonly use propofol. The result cannot prove that anaesthetic choice never modifies conditioning, because anaesthetic agents were not themselves randomised. It does show that removing most propofol exposure was insufficient to produce a demonstrable overall benefit. Hypotheses about other drugs, a fourth cycle or earlier treatment require prospective testing rather than retrospective use as reasons to disregard the result.

The lower-limb subgroup should be treated cautiously. Limb location was a prespecified subgroup in the main report, but choice of limb was not a separate randomised comparison. It may correlate with operation type, centre and patient characteristics. Numerous subgroup examinations increase the chance of an apparently unusual result, and the report does not provide convincing evidence of a reproducible interaction that should guide practice. The finding warrants investigation if the intervention is studied further; it does not justify recommending upper-limb conditioning as an effective alternative.

Readmission and petechiae also need different interpretations. Local petechiae have a direct and plausible connection with cuff inflation. Readmission encompasses diverse causes, including surgical problems, and was one of many adverse-event comparisons. Its interval barely excluded the null without multiplicity adjustment. The signal should remain visible, but it is insufficient to conclude that conditioning causes excess readmission. More fundamentally, an intervention with no established benefit does not require proof of major harm before routine adoption is rejected.

Finally, recruitment over nearly a decade and across many surgical procedures creates a broad but heterogeneous test. The trial is relevant to older patients receiving

antiplatelet therapy who undergo general anaesthesia, with less certainty for emergency presentations, unstable coronary disease or other conditioning schedules. Its principal contribution is a credible negative test of a feasible perioperative regimen. The next trial, if undertaken, needs a clearly justified population and outcome that can identify the proposed protective effect without assuming it exists.

Body of Evidence

ERICCA tested four upper-arm cycles after anaesthetic induction in 1,612 higher-risk patients undergoing coronary bypass surgery, with or without valve surgery. Cardiovascular death, nonfatal infarction, revascularisation or stroke at 12 months occurred in 26.5% with conditioning and 27.7% with sham, hazard ratio 0.95, 95% confidence interval 0.79–1.15. Biomarker release, kidney injury and recovery outcomes also showed no clear benefit. This was a strong clinical-outcome test in cardiac surgery, although the operative injury and anaesthetic environment differed from PRINCE.¹

RIPHeart, published in the same year, enrolled 1,403 cardiac-surgery patients and analysed 1,385. Under propofol-based anaesthesia, death, myocardial infarction, stroke or acute renal failure before discharge occurred in 14.3% with conditioning and 14.6% with sham. Troponin release and other recovery outcomes were also similar. The propofol setting became a proposed explanation for the lack of benefit. PRINCE's largely propofol-free design is therefore particularly relevant: it tested whether a major neutral result persisted when that potential interference was largely removed.²

PIXIE was more directly related to noncardiac surgery. It randomised 648 patients with cardiovascular risk factors undergoing emergency hip-fracture surgery, of whom 573 entered the reported analysis. Four upper-arm cycles reduced the revised myocardial-injury endpoint from 90/287 to 57/286, odds ratio 0.55, 95% confidence interval 0.37–0.80. Myocardial infarction was also less frequent, 10 versus 21 events. The endpoint had been revised because assay availability changed, and the trial could not establish an overall benefit in major cardiovascular events or mortality. Its older, predominantly female emergency hip-fracture population differed substantially from PRINCE's broader surgical cohort.³

The larger CONDI-2/ERIC-PPCI trial examined conditioning before primary coronary intervention for acute ST-elevation myocardial infarction. Of 5,401 randomised patients, 5,115 entered the principal analysis. Cardiac death or heart-failure admission at 12 months occurred in an estimated 9.4% with conditioning and 8.6% with control, hazard ratio 1.10, 95% confidence interval 0.91–1.32. This is indirect evidence for perioperative use, because it concerns established coronary occlusion and reperfusion. It nevertheless

shows that biological plausibility and earlier surrogate signals did not translate into clinical benefit in another large randomised setting.⁶

These studies should not be combined conceptually as though they tested identical interventions. They differ in three versus four cycles, timing relative to injury, anaesthesia, surgical stress, sham procedures and endpoint definitions. Such differences may explain heterogeneity, but they also limit the confidence with which a favourable small study can support broad implementation. A proposed mechanism is most useful when it generates a prospectively testable explanation for those differences.

The accompanying PRINCE editorial emphasised the limitations of a single postoperative troponin value and argued for more specific patient selection and clinical endpoints. Those are reasonable research considerations. They do not establish that a subgroup benefits, nor change the observed failure of this regimen to reduce the primary endpoint. The editorial's perspective and the trial's findings should remain distinct.⁷

By the end of 2025, PRINCE had added a substantial masked comparison in noncardiac surgery to a field dominated by heterogeneous studies. It did not confirm the expected cardioprotection, even with propofol largely avoided. Routine perioperative conditioning therefore remained unsupported; further investigation would require a more explicit account of who is expected to benefit, when the stimulus must be delivered and how that benefit will be measured.

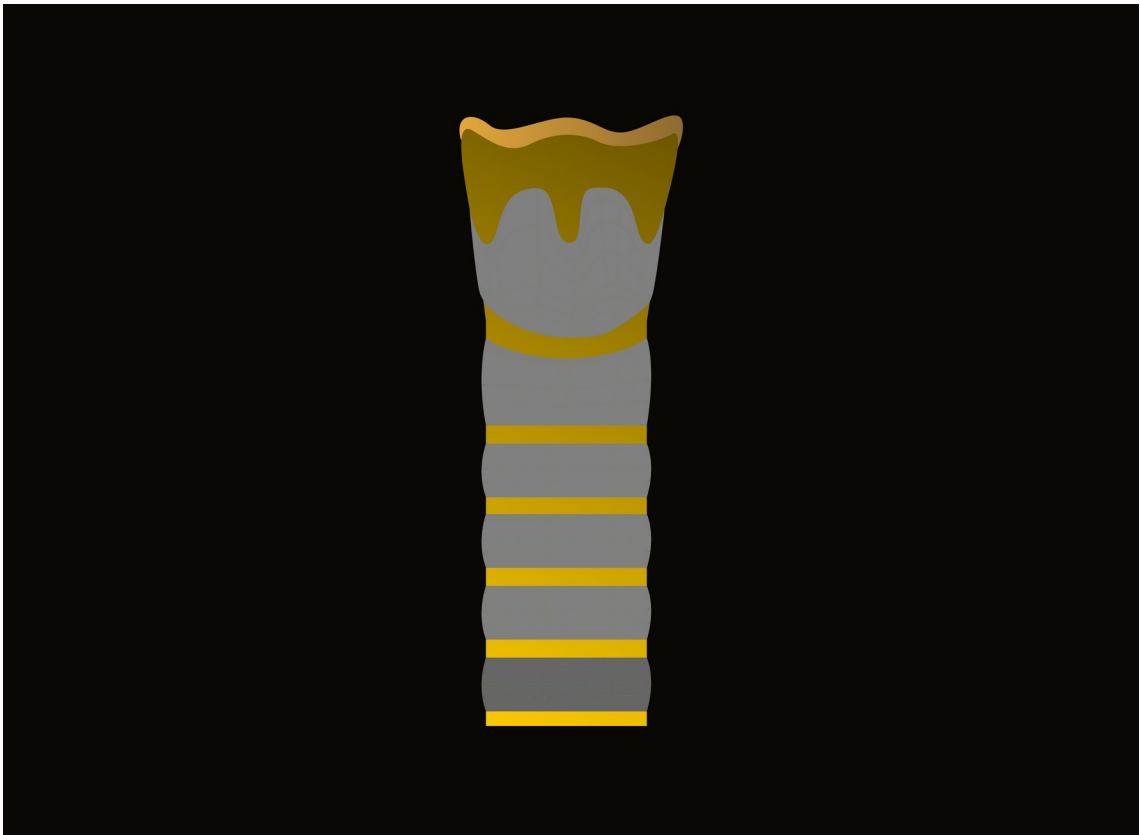
Should three cycles of remote ischaemic preconditioning be used routinely before noncardiac surgery?

No. PRINCE did not reduce postoperative myocardial injury or establish a benefit in clinical outcomes, despite largely avoiding propofol.

References

1. Hausenloy DJ, Candilio L, Evans R, Ariti C, Jenkins DP, Kolvekar S, et al. Remote Ischemic Preconditioning and Outcomes of Cardiac Surgery. *N Engl J Med*. 2015;373(15):1408-1417.
2. Meybohm P, Bein B, Brosteanu O, Cremer J, Gruenewald M, Stoppe C, et al. A Multicenter Trial of Remote Ischemic Preconditioning for Heart Surgery. *N Engl J Med*. 2015;373(15):1397-1407.

3. Ekeloef S, Homilius M, Stilling M, Ekeloef P, Koyuncu S, Münster AB, et al. The effect of remote ischaemic preconditioning on myocardial injury in emergency hip fracture surgery (PIXIE trial): phase II randomised clinical trial. *BMJ*. 2019;367:l6395.
4. Greco M, Lombardi G, Brusasco C, Pieri M, Roasio A, Monaco F, et al. Effect of Remote Ischemic Preconditioning on Myocardial Injury in Noncardiac Surgery: The PRINCE Randomized Clinical Trial. *Circulation*. 2025;152(17):1194-1205.
5. Greco M, Lombardi G, Brusasco C, Pieri M, Roasio A, Monaco F, et al. Supplementary methods and results to Effect of Remote Ischemic Preconditioning on Myocardial Injury in Noncardiac Surgery: The PRINCE Randomized Clinical Trial. *Circulation*. 2025.
6. Hausenloy DJ, Kharbanda RK, Møller UK, Ramlall M, Aarøe J, Butler R, et al. Effect of remote ischaemic conditioning on clinical outcomes in patients with acute myocardial infarction (CONDI-2/ERIC-PPCI): a single-blind randomised controlled trial. *Lancet*. 2019;394(10207):1415-1424.
7. Bøtker HE. Remote Ischemic Preconditioning in Noncardiac Surgery: Time to Reconsider or Refocus?. *Circulation*. 2025;152(17):1206-1208.



Airway Trials

Casey JD, Seitz KP, Driver BE, Gibbs KW, Ginde AA, Trent SA, et al. Ketamine or Etomidate for Tracheal Intubation of Critically Ill Adults. *N Engl J Med*. 2026;394(16):1608-1620. Published online December 9, 2025.

Introduction

Choosing an induction drug for emergency tracheal intubation requires balancing immediate cardiovascular stability against possible subsequent harm. Etomidate produces rapid hypnosis with relatively little haemodynamic disturbance, but inhibits adrenal steroid synthesis. A single dose can suppress cortisol production well beyond the brief period of anaesthesia. Whether this biochemical effect increases mortality, particularly in sepsis, has remained much less certain than the existence of adrenal suppression itself.

Ketamine avoids this adrenal effect and is often regarded as a cardiovascularly protective alternative. Its sympathomimetic actions, however, coexist with direct myocardial depressant and vasodilator effects. The net response in a critically ill patient cannot be assumed from the usual response in a physiologically stable person. The induction dose, preceding resuscitation, endogenous sympathetic activity and transition to positive-pressure ventilation may all influence what happens at the bedside. These physiological arguments explain the uncertainty; they do not establish which drug produces better outcomes.¹

Earlier randomised comparisons were inconclusive or produced different answers at different time points. A French trial demonstrated substantially more adrenal insufficiency with etomidate without a definitive difference in its primary organ-failure endpoint. The single-centre EvK trial subsequently favoured ketamine for seven-day survival but not for survival at day 28. A large contemporary randomised comparison was needed to determine whether avoiding etomidate actually saved lives.^{2,3}

The Randomized Trial of Sedative Choice for Intubation, RSI, addressed that clinical question in emergency departments and intensive care units. Its main hypothesis concerned mortality, while a prespecified procedural endpoint examined the immediate cardiovascular consequences of the induction choice. Keeping those two questions distinct is essential to understanding what the trial adds.¹

Synopsis

RSI was a pragmatic, multicentre, open-label, parallel-group randomised trial conducted at fourteen sites, comprising six emergency departments and eight intensive care units in six United States medical centres. Recruitment ran from 6 April 2022 to 10 August 2025. Funding included the Patient-Centered Outcomes Research Institute and the National Institutes of Health; funders had no role in trial conduct or analysis. An independent data and safety monitoring board oversaw the study. Emergency research provisions permitted enrolment without prospective informed consent. The principal report appeared online on 9 December 2025.^{1,4}

Adults undergoing tracheal intubation with an induction medication were eligible. Important exclusions included primary trauma, an immediate need for intubation that prevented randomisation, known pregnancy or prisoner status, and a clinician judgement that either study drug was required or contraindicated. Thus, the trial concerned patients for whom clinicians considered both options acceptable. It did not override a specific clinical contraindication or test either agent against no induction medication.

Allocation was 1:1, using variable-sized permuted blocks stratified by site. Sequentially numbered opaque envelopes concealed assignment until enrolment. Thereafter, clinicians and research personnel knew which drug had been allocated. A trained observer who was not performing the intubation recorded procedural physiology and treatments; subsequent clinical outcomes were obtained from records and follow-up sources.

Ketamine was administered intravenously, with a weight-based nomogram offering 2.0, 1.5 or 1.0 mg/kg. Etomidate was also intravenous, with corresponding options of 0.30, 0.25 or 0.20 mg/kg. Clinicians selected the dose within the assigned strategy. Other treatment decisions remained at their discretion, including pre-induction vasopressors, neuromuscular blockade, intravenous fluids, subsequent sedation and systemic corticosteroids. The intervention was therefore a choice of induction drug within usual resuscitation, rather than a fixed-dose comparison or an entire standardised airway bundle.

The primary endpoint was in-hospital death from any cause by day 28. Ascertainment for this endpoint ended at hospital discharge, so it was not identical to death in any location by day 28. The single prespecified secondary endpoint was cardiovascular collapse between induction and two minutes after intubation: systolic pressure below 65 mm Hg, new or increased vasopressor administration, or cardiac arrest. Exploratory outcomes included procedural blood pressure and oxygen saturation, first-attempt success, time to

intubation, and ventilator-, vasopressor- and ICU-free days. Prespecified safety outcomes included blood pressure and vasopressor receipt at 24 hours.

Assuming 30% mortality with etomidate, the investigators sought an absolute difference of 5.2 percentage points with 80% power and a two-sided significance level of 0.05. The enrolment target was 2,364. One planned interim analysis used a stringent stopping threshold of $P \leq 0.001$; the monitoring board recommended continuation without modification. The primary analysis used a generalised linear mixed-effects model with site as a random effect and treatment as a fixed effect. Additional baseline covariates were included in sensitivity analyses, rather than the main model. Secondary and exploratory confidence intervals were not adjusted for multiplicity.

Of 3,439 patients assessed, 2,367 underwent randomisation. Two subsequently identified prisoners were excluded, leaving 2,365 in the trial population: 1,176 assigned ketamine and 1,189 etomidate. Three patients in each group discontinued follow-up before primary-outcome ascertainment. Median age was 60 years, approximately 47% had sepsis or septic shock, and 22% were receiving vasopressors. Intubation occurred in an emergency department for 55.7% and in an ICU for 44.3%. The population therefore included substantial critical illness rather than predominantly uncomplicated airway procedures.

Treatment separation was excellent: 99.2% of the ketamine group received ketamine and 99.6% of the etomidate group received etomidate. Median doses were 1.6 and 0.28 mg/kg, respectively. Almost all patients received neuromuscular blockade. Video laryngoscopy was used for approximately 95%, and median systolic pressure at induction was 127 mm Hg in both groups. Additional induction sedatives were uncommon.

In-hospital death by day 28 occurred in 330/1,173 ketamine patients and 345/1,186 etomidate patients, 28.1% versus 29.1%. The site-adjusted risk difference was -0.8 percentage points (95% CI -4.5 to 2.9; $P=0.65$). Death in any location by day 28 occurred in 32.2% versus 32.4%, with an adjusted difference of 0.0 percentage points (95% CI -3.9 to 3.9). Other sensitivity analyses were consistent with the primary analysis. Among the 1,101 patients with sepsis or septic shock and primary-outcome data, mortality was 38.8% versus 38.2%; the adjusted difference was 1.0 percentage point (95% CI -4.8 to 6.7). Neither conventional subgroup analyses nor risk- and effect-modelling approaches identified convincing heterogeneity of the mortality effect.

Cardiovascular collapse occurred in 260/1,176 versus 202/1,189 patients, 22.1% versus 17.0%, an unadjusted difference of 5.1 percentage points (95% CI 1.9 to 8.3). New or increased vasopressor treatment occurred in 21.3% versus 15.9%. Systolic pressure

below 65 mm Hg occurred in 6.4% versus 5.5% of those with measurements, and cardiac arrest in 1.0% versus 0.8%; neither component individually established a clear difference. The composite difference was larger numerically among patients with sepsis and those with higher illness severity.

Other pressure measurements pointed in the same direction. Median lowest systolic pressure was 112 versus 118 mm Hg, and pressure below 80 mm Hg occurred in 14.4% versus 10.6%. First-attempt success and severe hypoxaemia were similar. Intubation took a median nine seconds longer with ketamine. Median ventilator-free days were 23 in both groups and vasopressor-free days 25 in both groups. At 24 hours, vasopressor receipt was 38.9% versus 42.3%, a difference of –3.4 percentage points (95% CI –7.5 to 0.7). Ventricular tachycardia, added as a post hoc outcome after the first 567 enrolments, occurred in 9/884 versus 2/905 patients.^{1,4}

Critique

RSI substantially weakens the argument that ketamine should routinely replace etomidate because avoidance of adrenal suppression improves survival. The trial found no such survival advantage, despite enrolling almost as many patients as the preceding randomised evidence combined. It also challenges the assumption that ketamine necessarily provides superior cardiovascular stability during emergency intubation. These are clinically consequential findings, but they are not proof that the drugs are interchangeable in every patient.

The mortality interval is more informative than the non-significant P value alone. The data are inconsistent with the large benefit sought in the sample-size calculation, yet still allow a ketamine benefit of several percentage points or a smaller disadvantage. A difference below the trial's detection target could matter for a frequently performed procedure. Calling the result equivalence would therefore overstate it. The defensible conclusion is that a clinically important proposed rationale for routinely favouring ketamine has not been demonstrated.

Mortality is objective and patient-centred, although relatively distant from a single induction dose. That distance is not a reason to dismiss the endpoint: the concern motivating the trial was delayed mortality caused by etomidate-associated adrenal suppression. The study appropriately tested the feared consequence rather than only measuring cortisol. Conversely, the result does not show that etomidate has no adrenal effect. It shows that the biochemical concern did not translate into a detectable net mortality disadvantage in this treatment setting.

The discharge-limited primary endpoint could have missed deaths after patients left hospital. The any-location sensitivity analysis is therefore important. Its almost identical event rates and balanced uncertainty make differential discharge an implausible explanation for a concealed large survival benefit. Missing primary outcomes were also very uncommon and equally distributed. These features give greater confidence in the mortality comparison than the headline primary analysis alone would provide.

The cardiovascular composite deserves a more qualified interpretation. Most events involved new or increased vasopressor treatment, a component influenced by clinical decisions in an unblinded trial. A clinician expecting instability after one drug could intervene at a different threshold. Moreover, a brief vasopressor adjustment and cardiac arrest do not have the same clinical importance. Describing the 5.1-percentage-point difference as an increase in cardiac arrest or irreversible circulatory failure would be incorrect.

However, the signal is not confined to that subjective component. Lower measured blood pressure, more readings below 80 mm Hg and more substantial pressure reductions all supported worse immediate haemodynamic tolerance with ketamine. These observations make it difficult to explain the entire composite difference solely by treatment expectations. The appropriate inference is concern about peri-intubation instability with the tested ketamine strategy, while recognising the limited precision of the most severe individual outcomes.

Multiplicity places another boundary around that conclusion. The report explicitly cautions that secondary and exploratory intervals were not adjusted for multiple comparisons. The prespecified cardiovascular endpoint has more credibility than an incidental finding selected from many measurements, but it does not acquire the same confirmatory status as a successfully tested primary endpoint. Ventricular tachycardia is more tentative still: ascertainment began after recruitment had started, the numbers were small, and the comparison was post hoc. It is a supporting safety observation rather than a definitive arrhythmia estimate.

The short procedural window is both useful and limiting. Measuring from induction to two minutes after intubation focuses on the immediate treatment decision, but cannot describe all haemodynamic consequences over subsequent hours. The 24-hour vasopressor comparison was numerically in the opposite direction and inconclusive. Those findings need not contradict each other: an agent can have different immediate and delayed effects, and survivors available for measurement at 24 hours form a selected population. One time point should not be used to erase the other.

Sepsis is particularly relevant because it has driven concern about etomidate. RSI included more than a thousand such patients, with no evidence of a mortality advantage from ketamine. This is stronger evidence than a small favourable subgroup in an earlier trial. Nevertheless, the sepsis interval remains wide enough to include meaningful effects, and numerical differences in cardiovascular collapse do not establish a proven sepsis-specific interaction. Neither the neutral overall result nor exploratory personalised-effect models identify a validated rule for selecting a drug for an individual patient.

The flexible dosing makes the result applicable to routine clinician-selected induction, but does not isolate the effect of dose. Median ketamine exposure was 1.6 mg/kg; the trial did not randomise a much smaller ketamine dose against etomidate. It would be reasonable to investigate whether another dose changes tolerance, but inappropriate to claim that such a regimen has already been shown safer. Conversely, the very high adherence means the main result is unlikely to be explained by substantial crossover or failure to deliver the allocated drug.

Applicability is strongest for non-trauma adults intubated in emergency departments or ICUs where both drugs are acceptable. Primary trauma, procedures too urgent for randomisation, pregnancy and cases with a specific drug indication were excluded. The high use of video laryngoscopy and contemporary airway support also matters when transferring findings to other environments. RSI neither compares propofol with these agents nor addresses prehospital behavioural sedation, prolonged ketamine infusions or continuous etomidate administration.

The practical judgement should therefore connect both outcomes. There is no demonstrated survival reward to offset ketamine's less favourable immediate haemodynamic profile in this broad comparison. Etomidate remains a reasonable induction option when otherwise appropriate. Ketamine also remains an available option for specific clinical circumstances, but a general claim that it is safer because it preserves adrenal function or always supports blood pressure is not supported by RSI.

Body of Evidence

The French randomised trial by Jabre and colleagues compared etomidate 0.3 mg/kg with ketamine 2 mg/kg in 655 acutely ill patients requiring emergency intubation. Its primary analysis included 469 patients after excluding those who died before hospital arrival or left intensive care within three days. The maximum early Sequential Organ Failure Assessment score was 10.3 versus 9.6, a mean difference of 0.7 (95% CI 0.0 to 1.4; $P=0.056$). Adrenal insufficiency was markedly more frequent with etomidate, with an odds ratio of 6.7. This established the physiological issue more clearly than a patient-centred advantage from avoiding it. The post-randomisation exclusions also mean its analysed population differed from RSI's much more complete mortality cohort.²

EvK subsequently randomised 801 patients at one United States centre to etomidate 0.2–0.3 mg/kg or ketamine 1–2 mg/kg. Among 791 analysed patients, seven-day survival was 77.3% with etomidate and 85.1% with ketamine, an absolute difference of 7.8 percentage points favouring ketamine. This was the prespecified primary endpoint. By day 28, survival was 64.1% versus 66.8%, with a difference compatible with either benefit or harm. EvK should not be relabelled a wholly negative trial, but its early result did not establish a persistent survival advantage. RSI directly tested whether a larger multicentre population reproduced a mortality benefit at the later time point.³

A separate single-centre randomised trial in 260 septic emergency-department patients also illustrates why physiological and survival findings should not be collapsed together. Day-28 survival was 80.8% with etomidate and 73.1% with ketamine, an inconclusive difference of 7.7 percentage points (95% CI –2.5 to 17.9). Vasopressor use within 24 hours was higher with etomidate, 43.9% versus 17.7%. That window differs from RSI's immediate procedural composite. A study of subsequent vasopressor receipt does not directly answer which drug causes more instability during induction, and neither trial established that its vasopressor findings mediated a survival effect.⁵

The preceding evidence therefore contained a consistent biological concern about adrenal suppression but an inconsistent clinical mortality signal. Differences in population, follow-up time, analysis exclusions and cardiovascular endpoint definitions make a single slogan about drug safety particularly unhelpful. RSI's large sample, excellent adherence, low missingness and inclusion of both emergency and intensive care settings provide an important corrective to strong conclusions drawn from smaller studies.

By the end of 2025, the direct evidence did not justify choosing ketamine routinely to improve survival over etomidate. The new trial also made immediate haemodynamic tolerance a more prominent consideration. Further work could reasonably examine

specific populations or dose strategies, but should preserve the separation between biochemical effects, transient treatment requirements and outcomes that matter to patients. None can automatically serve as a substitute for another.

Should ketamine routinely replace etomidate for emergency intubation to improve survival?

No. RSI found no survival advantage and raised concern about greater immediate haemodynamic instability with ketamine at the doses studied.

References

1. Casey JD, Seitz KP, Driver BE, Gibbs KW, Ginde AA, Trent SA, et al. Ketamine or Etomidate for Tracheal Intubation of Critically Ill Adults. *N Engl J Med*. 2026;394(16):1608-1620.
2. Jabre P, Combes X, Lapostolle F, Dhaouadi M, Ricard-Hibon A, Vivien B, et al. Etomidate versus ketamine for rapid sequence intubation in acutely ill patients: a multicentre randomised controlled trial. *Lancet*. 2009;374(9686):293-300.
3. Matchett G, Gasanova I, Riccio CA, Nasir D, Sunna MC, Bravenec BJ, et al. Etomidate versus ketamine for emergency endotracheal intubation: a randomized clinical trial. *Intensive Care Med*. 2022;48(1):78-91.
4. Casey JD, Seitz KP, Driver BE, Gibbs KW, Ginde AA, Trent SA, et al. Supplementary appendix to Ketamine or Etomidate for Tracheal Intubation of Critically Ill Adults. *N Engl J Med*. 2025.
5. Srivilaithon W, Bumrungphanithaworn A, Daorattanachai K, Limjindaporn C, Amnuaypattanapon K, Imsuwan I, et al. Clinical outcomes after a single induction dose of etomidate versus ketamine for emergency department sepsis intubation: a randomized controlled trial. *Sci Rep*. 2023;13(1):6362.

PreVent 2

Treggiari MM, Sharp ES, Ohnuma T, Fajardo EC, Aydin A, Akhtar S, et al. Hospital and long-term outcomes for subglottic suction and polyurethane cuff versus standard endotracheal tubes in emergency intubation (PreVent 2): a randomised controlled phase 2 trial. Lancet Respir Med. 2026;14:141-150. Published online November 27, 2025.

Introduction

An endotracheal tube provides an essential airway while also creating a route through which contaminated secretions can reach the lungs. Secretions collect above its cuff, and folds in cuff material can permit leakage despite an apparently adequate seal. A thinner polyurethane cuff and a separate channel for drainage above the cuff offer plausible ways to reduce this exposure. Their value, however, depends on whether improved secretion management prevents clinically important infection without causing additional airway injury.

Subglottic secretion drainage has often been assessed through rates of ventilator-associated pneumonia. That endpoint is difficult to interpret because clinical signs, radiographic abnormalities and respiratory cultures do not reliably distinguish invasive infection from other causes of respiratory deterioration or colonisation. An intervention that changes secretions could also change the likelihood of meeting a pneumonia definition without improving recovery. Ventilator duration, antimicrobial exposure, survival and longer-term health therefore provide useful complementary evidence.¹

The combined polyurethane-cuff and suction design introduces a further trade-off. Continuous suction might injure mucosa, while a more complex tube costs more and requires additional bedside work. Earlier studies suggested fewer pneumonias in selected patients expected to require prolonged ventilation, but the balance of benefit and harm from routinely using these tubes during emergency intubation remained uncertain. The link between pneumonia prevention and better cognition or quality of life months later was even less established.²

PreVent 2 evaluated the combined device in a broad emergency-intubation population. Unusually, its coprimary outcomes concerned six-month airway symptoms, quality of life and cognition, with hospital ventilator-associated events assessed as secondary

outcomes. This ambitious patient-centred design makes the completeness of long-term follow-up central to its interpretation.

Synopsis

PreVent 2 was a randomised phase 2 trial conducted sequentially at two United States hospitals. Oregon Health and Science University enrolled patients from 6 May to 31 July 2019; Yale New Haven Hospital recruited from 29 September 2020 to 11 February 2022. The National Institutes of Health and National Heart, Lung, and Blood Institute funded the study without involvement in its conduct or interpretation. Emergency research provisions permitted initial enrolment without prospective consent, followed by an approach to patients or representatives for continued participation.^{2,3}

Adults requiring emergency tracheal intubation in the emergency department or elsewhere in hospital for acute respiratory distress or failure were eligible. Elective operating-room intubation was excluded, as were protected populations and patients who had declined research participation. There was no requirement to predict a prolonged period of ventilation. This strategy tested routine initial tube selection in a broadly eligible emergency population.

A computer-generated fair-coin allocation sequence assigned patients equally in expectation to conventional polyvinylchloride-cuffed tubes or polyurethane-cuffed tubes with subglottic suction, abbreviated PU-EVAC. Tubes were supplied in opaque sealed intubation kits, concealing assignment until use, in internal diameters of 7.0, 7.5 and 8.0 mm. Staff replenished kits daily. Although allocation was concealed and cognitive assessors were masked, the suction apparatus was apparent during subsequent clinical care. The trial should not be understood as having completely masked bedside treatment throughout admission.

Patients allocated PU-EVAC received continuous suction through the subglottic port at –20 mm Hg until tube removal. Small manual air boluses were used to clear occlusions and routinely every four hours. Conventional tubes received usual care without subglottic drainage. Both designs had high-volume, low-pressure cuffs. The report specifies a cuff-pressure target of 25 mm Hg, measured with a manometer at least twice daily; that stated unit should be distinguished from the cm H₂O commonly used in clinical cuff-pressure protocols. A replacement tube matching the original model and size was kept at the bedside to limit crossover.

The coprimary outcomes at six months were laryngeal injury, health-related quality of life and cognitive function. Laryngeal injury was assessed with a symptom questionnaire

covering voice, throat, swallowing and respiratory complaints, rather than systematic endoscopic examination. Quality of life was measured with the SF-36 physical and mental component summary scores. A trained, masked assessor administered a battery of cognitive tests, with performance compared against normative data and grouped into domains. Remote assessments were introduced when in-person visits were impractical. The investigators did not specify a single joint success rule or multiplicity-adjusted decision rule for the coprimary endpoints.

Hospital outcomes included Centers for Disease Control surveillance categories. A ventilator-associated condition required a sustained rise in minimum oxygen concentration or positive end-expiratory pressure after a stable baseline. Infection-related ventilator-associated complications, IVACs, additionally required inflammatory criteria and a new antibiotic course lasting at least four days. Possible ventilator-associated pneumonia added microbiological evidence. These algorithmic endpoints differ from all clinically suspected pneumonias. Other outcomes included antibiotics, ventilator-free days to day 28, hospital stay, mortality, tracheostomy and organ support.

The planned sample of 1,074 provided 80% power for a ten-percentage-point difference in laryngeal injury under the stated assumptions. The calculations also considered quality-of-life differences and an anticipated control IVAC rate of approximately 11%, with a large relative reduction sought. Complete-case analyses followed original assignment, using regression with robust standard errors. Multiple imputation explored missing-data sensitivity, and per-protocol and site-adjusted analyses were supplied. No non-inferiority or equivalence margin was tested.

Of 1,074 initially randomised patients, six were excluded because of inadvertent enrolment or refusal, leaving 535 PU-EVAC and 533 conventional-tube patients. Only the first 90 enrolments occurred at Oregon; the remainder were at Yale. Mean age was 62.9 years and 63% were male. Comorbidity was substantial, and roughly a quarter had COVID-19. Seventy-nine patients underwent re-randomisation after reintubation; 515 PU-EVAC and 510 conventional-tube patients never crossed over. The principal analyses retained the initial allocation.

By six months, 558 patients had died: 274/535 in the PU-EVAC group and 284/533 controls, 51% versus 53%. A further 167 did not consent to continued participation and 186 were lost to follow-up. Only 157 completed the follow-up assessment, 87 versus 70. Individual endpoints had slightly smaller denominators. Thus, six-month functional data were available for about 15% of the original cohort and 31% of those known to survive.

Laryngeal symptoms were identified in 71/86 PU-EVAC patients and 49/70 controls, 83% versus 70%. The main report gives a Rao–Blackwell-adjusted risk difference of 0.11 (95% CI –0.03 to 0.20; $P=0.098$). Physical component scores averaged 39.97 versus 40.49, a difference of –0.52 (95% CI –4.21 to 3.17). Mental component scores were 45.91 versus 48.89, a difference of –2.98 (95% CI –7.03 to 1.07). Cognitive impairment was present in 70/82 versus 54/67 patients, 85% versus 81%, with a risk difference of 0.05 (95% CI approximately –0.07 to 0.17). None of the principal comparisons established a benefit.

IVAC occurred in 43/535 versus 33/533 patients, 8% versus 6%, and possible pneumonia in 30/535 versus 24/533, 6% versus 5%. The reported differences were approximately two percentage points (95% CI –1 to 5) and one percentage point (95% CI –2 to 4), respectively. In-hospital mortality was 41% versus 45%, an inconclusive difference. Antibiotic exposure, ventilator-free days and other organ-support outcomes did not demonstrate improvement. Mean hospital stay was four days longer with PU-EVAC (95% CI 0.1 to 7.9), an isolated secondary result. Complete-case, imputed and other sensitivity analyses broadly supported the absence of demonstrated benefit, although several supplementary estimates require editorial clarification.^{2,3}

Critique

PreVent 2 provides no affirmative evidence for routinely selecting this combined tube design during emergency intubation. Hospital infection-related events were not reduced, and no consistent improvement appeared in outcomes beyond those surveillance definitions. Its long-term results are much less definitive: extensive mortality and loss of follow-up leave a small selected group in whom absence of a significant difference cannot establish equivalent safety, cognition or quality of life.

The trial asks a useful implementation question. At emergency intubation, clinicians cannot always predict who will require prolonged ventilation. Testing an unselected eligible population therefore reflects a possible routine procurement policy. The same design reduces direct comparability with studies deliberately enriching for prolonged ventilation and higher pneumonia risk. A neutral result in this broad population does not prove that every patient at particularly high risk receives no benefit, but it places the burden of demonstrating such a benefit on evidence specific to that subgroup.

The observed control IVAC rate was substantially below the planning assumption, leaving fewer preventable events and less power for the large reduction anticipated. This limits sensitivity to modest benefit. It does not explain away the direction of the findings: both IVAC and possible pneumonia were numerically more common with PU-EVAC, and associated measures such as antibiotic use and ventilator-free days gave

no coherent indication of improved recovery. The study is therefore more than an isolated non-significant pneumonia comparison, even though its confidence intervals still allow some benefit or harm.

Surveillance definitions improve reproducibility but change the clinical question. An IVAC requires deterioration in ventilator settings after stability and a particular antibiotic pattern. It will not identify every clinically diagnosed pneumonia, and some qualifying deterioration can have non-infectious contributions. Conversely, more subjective pneumonia definitions can be influenced by changes in secretion burden or culture positivity. Neither category is a perfect biological diagnosis. The convergence of surveillance, treatment and recovery outcomes is more persuasive than declaring one definition universally correct.

The largest limitation concerns the coprimary endpoints. Death precludes assessment of later cognition or symptoms, while refusal and loss of contact create additional missingness among survivors. Randomisation does not guarantee that the survivors who agree to extensive testing remain comparable or representative. Similar mortality and broadly similar loss rates between arms are reassuring only to a point: equal proportions missing can still conceal different patterns of health and recovery. A complete-case comparison estimates outcomes among the observed participants, not an uncomplicated treatment effect on six-month health in everyone randomised.

Multiple imputation cannot recreate information without assumptions. Its value depends on whether recorded variables adequately account for why outcomes are missing. Here, prolonged illness, disability, cognitive impairment or recovery itself could affect willingness and ability to attend. The published sensitivity analyses assume an ignorable missing-data mechanism; they do not exclude bias when unobserved outcomes influence participation. Death also cannot be handled as though it were merely an omitted SF-36 questionnaire. The findings should therefore not be promoted as evidence that the device definitively fails to affect long-term recovery.

The airway endpoint also needs careful naming. The reported “laryngeal injury” was mainly a broad symptom assessment without systematic specialist examination. Dyspnoea, cough or swallowing difficulty can arise from many sequelae of critical illness, and are not proof of a structural lesion caused by suction. The higher symptom rate is a safety concern deserving further investigation, but it does not establish a definite increase in anatomically confirmed laryngeal damage. Similarly, poor cognitive scores relative to population norms do not show that intubation or the tube caused new dementia; baseline cognition and survival selection matter.

The safety results contain reporting inconsistencies that warrant resolution before publication. The primary text, supplementary overall symptom row and unadjusted estimate do not present an entirely consistent adjusted effect. Some supplementary confidence intervals and P values are also difficult to reconcile. The chapter retains the main report's designated estimate without silently reconstructing an alternative. These discrepancies do not reveal a hidden benefit, but they limit precise quantification of the possible airway harm.

Allocation concealment should not be confused with full clinical masking. Opaque kits protect the randomisation decision, whereas the visible drainage port and active suction reveal the intervention during care. Masked neuropsychological assessment is a separate strength. Hospital decisions, including antibiotics, can still be influenced by knowledge of treatment. In this trial, the lack of benefit across several distinct outcomes reduces reliance on any one potentially influenced measurement.

Reintubation and re-randomisation complicate exposure. Initial allocation remains the appropriate starting point for a policy analysis, but some patients experienced more than one device strategy. A per-protocol analysis can explore this dilution while introducing selection by post-randomisation events. Its nominally stronger symptom signal is not a substitute for the primary comparison, particularly given small numbers and inconsistencies in supplementary reporting. The study also combines two modifications, so it cannot separately identify the effects of polyurethane material, continuous suction or their interaction.

Finally, most recruitment occurred at one hospital during the pandemic, with substantial comorbidity and high mortality. Calling it multicentre should not imply a broadly balanced international sample. The four-day hospital-stay difference is exploratory and should not dominate interpretation in the presence of many comparisons and competing outcomes. The most supportable practical position is to avoid assuming that this more elaborate tube routinely improves emergency-intubation outcomes, while retaining uncertainty about selected high-risk populations and long-term airway effects.

Body of Evidence

Earlier trials generally placed pneumonia prevention at the centre of the question. The PreVent 2 report describes a 280-patient randomised study in patients expected to need ventilation for more than 24 hours, in which the combined polyurethane-cuff and drainage design reduced diagnosed pneumonia from approximately 22% to 8%. That was a clinically selected population with a substantially higher event rate than PreVent 2. The difference in baseline risk and pneumonia definition matters when considering whether those results should support routine use for every emergency intubation. The original abstract confirms 31/140 versus 11/140 pneumonias.⁴

The preceding PreVent pilot informed feasibility and sample-size assumptions, including the unexpectedly important question of later airway symptoms. Moving from a pilot to a large emergency-enrolment programme increased the number exposed to the device but did not ensure successful long-term assessment. PreVent 2 illustrates that an acute intervention can be easy to deliver within routine care while demanding follow-up tests remain difficult to obtain. The two stages answer different aspects of whether such research can determine patient-centred benefit.²

Caroff and colleagues' systematic review included seventeen randomised trials and 3,369 patients. Subglottic drainage reduced diagnosed pneumonia (risk ratio 0.58, 95% CI 0.51–0.67), but did not clearly improve ventilation duration, ICU or hospital stay, mortality, or antibiotic use. A separate meta-analysis of polyurethane cuffs alone, discussed in the principal report, gave an imprecise estimate.⁵ These interventions should not be treated as identical. Continuous versus intermittent suction, cuff design, expected ventilation duration and baseline prevention practices can all change the comparison. Furthermore, a pooled reduction in a diagnostically unstable endpoint is more convincing if accompanied by fewer ventilator days, less antibiotic exposure or better survival.^{1,2}

The accompanying editorial by Klompas and Branson places PreVent 2 within the evolution of prevention guidance. They describe how the 2022 guidance moved subglottic drainage from an essential practice to an optional additional approach after updated syntheses failed to demonstrate consistent benefits in ventilation duration, length of stay or mortality. Their concern is that reducing secretions or colonisation can reduce the probability of a pneumonia diagnosis without necessarily reducing invasive infection. This is an interpretation of the evidence and the rationale for the guidance, rather than an additional randomised result.¹

PreVent 2 strengthens the case for requiring outcomes beyond diagnosed pneumonia before universal adoption. It tested the combined device from the initial emergency

procedure, monitored adherence and assessed hospital events in a comparatively large randomised cohort. No corresponding reduction in infection-related complications, antibiotic use or ventilation burden emerged. At the same time, its ambitious six-month assessment cannot resolve the long-term question because so few survivors contributed measurements.

The resulting evidence supports restraint in routine use rather than a claim that all subglottic drainage has been disproved. A future study of a clearly defined high-risk population would need reliable device delivery, clinically meaningful infection and recovery outcomes, and a workable plan for airway assessment and survivor follow-up. Until then, neither physiological plausibility nor older reductions in pneumonia diagnoses establish a patient-centred advantage for routine PU-EVAC tube selection.

Should polyurethane-cuffed tubes with continuous subglottic suction be used routinely for emergency intubation?

PreVent 2 does not support routine adoption. It found no hospital-outcome benefit, while extensive missing follow-up leaves long-term benefit and airway harm uncertain.

References

1. Klompas M, Branson R. Time to pull the plug on subglottic secretion drainage? *Lancet Respir Med*. 2026;14:104-105. Published online November 27, 2025.
2. Treggiari MM, Sharp ES, Ohnuma T, Fajardo EC, Aydin A, Akhtar S, et al. Hospital and long-term outcomes for subglottic suction and polyurethane cuff versus standard endotracheal tubes in emergency intubation (PreVent 2): a randomised controlled phase 2 trial. *Lancet Respir Med*. 2026;14:141-150. Published online November 27, 2025.
3. Treggiari MM, Sharp ES, Ohnuma T, Fajardo EC, Aydin A, Akhtar S, et al. Supplementary appendix, results pages 1-9, to Hospital and long-term outcomes for subglottic suction and polyurethane cuff versus standard endotracheal tubes in emergency intubation (PreVent 2). *Lancet Respir Med*. 2025.
4. Lorente L, Lecuona M, Jiménez A, Mora ML, Sierra A. Influence of an endotracheal tube with polyurethane cuff and subglottic secretion drainage on pneumonia. *Am J Respir Crit Care Med*. 2007;176(11):1079-1083.
5. Caroff DA, Li L, Muscedere J, Klompas M. Subglottic Secretion Drainage and Objective Outcomes: A Systematic Review and Meta-Analysis. *Crit Care Med*. 2016;44(4):830-840.



Respiratory Trials

SESAR

Jabaudon M, Quenot JP, Badie J, Audard J, Jaber S, Rieu B, et al. Inhaled Sedation in Acute Respiratory Distress Syndrome: The SESAR Randomized Clinical Trial. JAMA. 2025;333(18):1608-1617.

Introduction

Sedation in acute respiratory distress syndrome (ARDS) must reconcile several competing objectives. Patients need relief from distress, sufficient tolerance of invasive ventilation and, at times, immobility for prone positioning or neuromuscular blockade. Excessive or prolonged sedation can obstruct liberation from ventilation and neurological assessment. The choice of sedative therefore represents an intervention within a wider treatment strategy, with potential effects on circulation, respiratory mechanics and recovery as well as consciousness.

Volatile anaesthetics offer an attractive alternative to intravenous sedatives. Rapid pulmonary elimination might permit predictable awakening, while experimental anti-inflammatory and epithelial-protective effects suggest benefits extending beyond sedation. However, these physiological advantages require clinical confirmation. Delivery in an intensive care unit (ICU) introduces an anaesthetic-conserving device into the ventilator circuit, and prolonged exposure differs substantially from the shorter administration usual during surgery. The pharmacology and the delivery system both require evaluation in injured lungs.

A small randomised study comparing sevoflurane with midazolam in early ARDS found better oxygenation after 48 hours and changes in inflammatory and epithelial-injury biomarkers. A larger isoflurane study subsequently demonstrated effective sedation compared with propofol in a broader mechanically ventilated population. These findings supported further investigation but did not establish that sevoflurane improves survival or accelerates recovery from ARDS. They involved different drugs, comparators, durations and principal outcomes.^{1,2}

SESAR tested the clinically consequential next step: whether an early strategy using inhaled sevoflurane, rather than intravenous propofol, increased days alive and free of invasive ventilation in moderate to severe ARDS. The trial was designed to examine an established syndrome within a defined hypoxaemia threshold, rather than to demonstrate that an inhaled agent could produce an acceptable sedation score.³

Synopsis

SESAR was an investigator-initiated, multicentre, open-label, assessor-blinded, superiority trial conducted in 37 French ICUs from May 2020 to October 2023. The University Hospital of Clermont-Ferrand sponsored the study. Funding came from the French Ministry of Health, the European Society of Anaesthesiology and Intensive Care, and Sedana Medical. The company supplied sevoflurane and delivery equipment and supported training; the supplementary report states that funders had no role in study design, data analysis or writing. Independent safety oversight included regular reports and a planned interim analysis.^{3,4}

Adults receiving invasive ventilation were eligible within 24 hours of meeting ARDS criteria, with a $\text{PaO}_2:\text{FiO}_2$ ratio below 150 mm Hg and positive end-expiratory pressure (PEEP) of at least 8 cm H_2O . Important exclusions included extracorporeal membrane oxygenation at enrolment, prolonged pre-existing invasive ventilation, chronic hypercapnic respiratory failure, body mass index above 40 kg/m², suspected intracranial hypertension, and contraindications to the study drugs. Patients expected to die within 24 hours or require ventilation for less than 48 hours were excluded. Eligibility also incorporated the delivery device's minimum tidal-volume requirement.

Allocation was performed centrally online in a 1:1 ratio, using minimisation and stratification by site, severity of hypoxaemia, suspected or confirmed COVID-19, and shock. The clinical teams could not be blinded to the route of administration. Outcomes were assessed with masking to allocation. Emergency consent arrangements permitted enrolment before patients recovered capacity, with subsequent information and consent procedures. Patients withdrawing consent, or placed under judicial protection, were excluded from the analysis.

Sevoflurane was delivered through a Sedaconda ACD-S device between the endotracheal tube and ventilator circuit, with gas scavenging and bedside monitoring of inspired and expired concentrations. The comparator was intravenous propofol. Both strategies continued until sedation was no longer required or seven days had elapsed; sedation restarted within this period used the assigned strategy. Analgesia followed an analgesia-first approach, and additional sedatives could be used. Thereafter, sedation was determined by the treating clinicians.

Both groups initially targeted deep sedation, with a Richmond Agitation-Sedation Scale (RASS) score of -5 to -4, and cisatracurium for up to 48 hours. Earlier relaxation of this approach was permitted after sustained improvement in oxygenation. Subsequently, lighter sedation, targeting RASS -1 to 0, daily awakening and spontaneous-breathing trials were recommended. Neuromuscular blockade beyond 48 hours was treated as

rescue therapy. Low tidal volumes and a high-PEEP strategy were shared co-interventions; respiratory rescue treatments remained clinician-directed. The trial recommended a stabilisation period before prone positioning.⁴

The primary endpoint was ventilator-free days from randomisation to day 28. Successful unassisted breathing had to persist for at least two consecutive calendar days and through day 28. After reintubation, counting restarted at the end of the final period of assisted breathing. Death before day 28, or continued assisted breathing at day 27, resulted in zero ventilator-free days. The key secondary endpoint was 90-day survival. Other outcomes included earlier mortality and ICU-free days. Safety assessment covered haemodynamics, kidney injury and predefined adverse events through day seven.^{3,4}

The sample-size calculation assumed approximately 30-35% mortality by day 28 and a standard deviation of eight ventilator-free days. A sample of 680 would provide more than 80% power to detect a mean difference of two days; the recruitment target was 700. After a blinded interim analysis at 350 participants, the monitoring committee recommended continuation. The principal analysis used a modified intention-to-treat population and a mixture of generalised gamma distributions modelling successful weaning and competing death. Results were expressed as a between-group difference and a standardised hazard ratio. Survival was examined using Kaplan-Meier methods and Cox regression. Adjusted, per-protocol and subgroup analyses were also undertaken. Confidence intervals were not adjusted for multiplicity.

Of 2,016 screened patients, 700 were randomised, 350 to each group. Thirteen were subsequently excluded: four receiving sevoflurane allocation and nine receiving propofol allocation. The analysis therefore included 346 and 341 patients, respectively. Mean age was approximately 65 years; about 30% were women. Pneumonia predominated, 54.9% had COVID-19 pneumonia, and approximately three-quarters were receiving vasopressor or inotropic support before randomisation. Median baseline $\text{PaO}_2\text{:FiO}_2$ ratios were 111 and 107 mm Hg. Most patients had already received sedation, usually propofol.

One patient in each group did not receive the assigned drug. Median study-drug exposure was five days with sevoflurane and six days with propofol; total sedation duration was seven days in both groups. Deep sedation was achieved during the first 48 hours in 93.9% and 89.2%, respectively. The principal report records additional sedation in 148 sevoflurane patients and 64 propofol patients. Continuous neuromuscular blockade was used during the first 48 hours in 72.1% and 70.0%; its total median duration was five days in both groups. Thus, actual sedation and paralysis commonly extended beyond the intended initial period.

Median ventilator-free days were zero in both groups, with interquartile ranges of 0-11.9 days for sevoflurane and 0-18.7 for propofol. The reported between-group estimate was -2.1 days (95% confidence interval [CI] -3.6 to -0.7), with a standardised hazard ratio of 0.76 (95% CI 0.50-0.97). These are the study's reported inferential estimates; the -2.1-day value is not subtraction of the two observed group medians. Primary-outcome subgroup analyses did not identify convincing treatment-effect heterogeneity.^{3,4}

By day 90, 183/346 patients receiving sevoflurane and 151/341 receiving propofol had died: 52.9% versus 44.3%, an absolute difference of 8.6 percentage points (95% CI 1.2-16.1). The mortality hazard ratio was 1.31 (95% CI 1.05-1.62), with log-rank P=0.02. At day 28, mortality was 152/345 versus 132/340, or 44.1% versus 38.8% (risk ratio 1.13, 95% CI 0.95-1.36). ICU-free days also favoured propofol. Adjusted analyses supported the main direction of effect, although some per-protocol and sensitivity estimates were less precise.

Fewer sevoflurane patients remained free of acute kidney injury through day seven: 110/346 (31.8%) versus 146/341 (42.8%). Norepinephrine requirements were higher, and five cases of nephrogenic diabetes insipidus occurred with sevoflurane versus none with propofol. Severe hypercapnic acidosis occurred in 11 versus five patients. Two events classified as malignant hyperthermia occurred in the sevoflurane group, while three classified as propofol-related infusion syndrome occurred in the propofol group. These uncommon events had limited precision. Oxygenation did not show the expected advantage with sevoflurane; carbon dioxide levels and ventilatory ratio were higher during much of the intervention period.^{3,4}

Critique

SESAR provides evidence against adopting this sevoflurane strategy for early moderate to severe ARDS. The primary outcome favoured propofol, and the mortality signal pointed in the same direction. This combination is more concerning than an isolated adverse laboratory finding or failure to demonstrate a hypothesised physiological benefit. Nevertheless, the conclusion must remain attached to the comparison actually tested: sevoflurane delivered through an anaesthetic-conserving device for up to seven days within a strategy that initially used deep sedation and often prolonged neuromuscular blockade.

The primary endpoint requires careful explanation. Assigning zero days to death appropriately prevents early death from appearing to shorten ventilation. It also creates a large concentration of values at zero, combining patients who died with survivors who remained ventilated. The zero medians are consequently an incomplete description of

the distribution. The reported -2.1-day contrast must not be described as the literal difference between those medians, nor as two extra ventilation days in every survivor. Its practical interpretation is supported by the direction of the full outcome distribution and by examination of mortality, rather than by the descriptive medians alone.

Death also complicates interpretation of ventilation duration considered separately. The supplement reports median durations of invasive ventilation of 12.7 and 11.1 days, without a convincing difference. This does not contradict the primary finding: a treatment associated with more deaths can truncate time receiving ventilation. Conversely, the day-90 mortality result extends beyond the primary endpoint's 28-day window and supplies clinically important information that ventilator-free days cannot contain. Mortality is less susceptible than extubation timing to bedside judgement, although decisions about limitations of treatment remain part of unblinded critical care.

The 8.6-percentage-point excess in day-90 mortality is substantial. Its interval ranges from a relatively small excess to a very large one, so the magnitude remains uncertain. Survival was the designated key secondary outcome, but the report states that intervals were not adjusted for multiplicity; the P value should therefore not be treated as a separately multiplicity-protected confirmation of mortality harm. Equally, uncertainty about formal secondary testing does not make this signal clinically ignorable. An intervention with an adverse primary outcome, no established compensating benefit and concerning survival does not warrant routine adoption while awaiting a more definitive mechanism.

Treatment separation was meaningful but not pharmacologically pure. Patients assigned to sevoflurane frequently received additional intravenous sedation, and study-drug exposure was somewhat shorter than with propofol. This limits attribution to sevoflurane alone. However, extra sedation may represent difficulty delivering the assigned strategy and therefore belongs within its clinical consequences. Excluding patients who needed treatment changes would not necessarily reveal an unbiased biological effect: the reasons for switching may themselves be associated with illness severity or response to treatment. The modified intention-to-treat comparison should retain primacy over the selected per-protocol populations.

Deep sedation and prolonged paralysis materially constrain applicability. Both groups were treated similarly in principle, so these shared practices do not by themselves explain away the randomised difference. They may, however, modify how a drug or device behaves, and they distinguish SESAR from a unit that achieves early light sedation with little neuromuscular blockade. Median paralysis of five days and very limited delirium assessment show that this was not chiefly a rapid-awakening trial. The report

cannot establish either neurological benefit or equivalence between the sedatives in a lighter, shorter sedation strategy.

The device is a plausible contributor to the adverse respiratory pattern. It adds apparatus dead space, while the trial observed higher PaCO₂, lower pH and higher ventilatory ratio despite generally protective tidal volumes. Increasing respiratory rate to compensate might impose additional mechanical load. These observations justify concern, but they do not prove a device-mediated causal pathway. Comparator humidification and apparatus dead space were not recorded sufficiently to isolate that explanation, and the post hoc mechanical-power surrogate does not substitute for a randomised comparison of devices. The defensible conclusion concerns the tested drug-and-delivery strategy.

Renal and haemodynamic findings strengthen concern without identifying a single cause. Higher vasopressor requirements, fewer patients free of kidney injury and nephrogenic diabetes insipidus are compatible with clinically important adverse effects. Fluoride-mediated toxicity and vasodilatation are possible explanations, but fluoride concentrations were not measured, and fluid volumes, opioid exposure and expired sevoflurane concentrations were not collected. It would be inappropriate to state that a specific toxicity caused the deaths. These gaps matter mainly to understanding and modifying the intervention, rather than to deciding whether the tested strategy should become standard care.

The COVID-19 subgroup deserves restraint. Ninety-day mortality differed little between treatments among patients with COVID-19 but separated more among those without it; the reported interaction P value was 0.01. In contrast, the interaction for the primary endpoint was not convincing. Multiple subgroup assessments, different baseline risks and limited precision prevent a reliable claim that sevoflurane is safe or beneficial in COVID-19 ARDS. Importantly, the trial's pandemic case mix does not support dismissing the overall finding as uniquely caused by COVID-19 when the larger mortality separation occurred outside that subgroup.

There are genuine protections against several alternative explanations. Allocation used a central system, baseline respiratory severity and circulatory support were reasonably similar, and the study incorporated multiple ICUs and blinded outcome assessment. Excluding the first five patients at each centre did not reverse the primary direction, weakening a simple explanation based entirely on initial unfamiliarity. However, some supporting intervals crossed the null. Describing every sensitivity analysis as independently statistically significant would overstate robustness. Thirteen post-randomisation exclusions were few but should remain visible, and the distinction between 700 randomised and 687 analysed should be preserved.

Overall, uncertainty about mechanism and extrapolation is different from uncertainty about whether to introduce this particular practice. SESAR does not condemn brief sevoflurane anaesthesia, establish a class effect for all volatile agents, or test isoflurane against propofol in a lightly sedated general ICU population. It does substantially weaken the case for prolonged sevoflurane sedation as an ARDS treatment and provides a clear reason to prefer alternatives in patients resembling those enrolled.

Body of Evidence

The initial clinical rationale came from Jabaudon and colleagues' 50-patient pilot study. Adults within 24 hours of moderate to severe ARDS were randomised to sevoflurane or midazolam for 48 hours. The primary endpoint was oxygenation on day two. Mean $\text{PaO}_2:\text{FiO}_2$ was 205 versus 166 mm Hg, with $P=0.04$, and selected inflammatory and epithelial-injury markers also favoured sevoflurane. This was useful evidence of physiological activity, but the small trial was not designed to establish a survival advantage. Its comparator was midazolam, not propofol, and exposure was substantially shorter than in SESAR.¹

The Sedaconda trial addressed a different question: whether isoflurane could provide sedation at least as effectively as propofol in invasively ventilated ICU patients. It enrolled 301 patients and evaluated a much shorter treatment period than SESAR. Demonstrating time within a target sedation range, and favourable awakening characteristics, establishes practical sedative efficacy. It does not demonstrate that an inhaled agent improves the course of ARDS. The distinction between isoflurane and sevoflurane also matters; efficacy and safety should not be transferred between agents solely because both are volatile anaesthetics.²

An a priori hypoxaemic-respiratory-failure substudy of Sedaconda included 162 participants with baseline $\text{PaO}_2:\text{FiO}_2$ at or below 300 mm Hg. Isoflurane did not produce greater improvement in oxygenation than propofol over the treatment period. The subgroup with more severe hypoxaemia likewise did not establish an oxygenation benefit from isoflurane. No convincing advantage in mortality or ventilator-free days was identified. These patients were drawn from the parent trial and must not be counted as an independent randomised replication. The findings nonetheless weaken the assumption that a volatile agent necessarily improves gas exchange compared with propofol.⁵

The Swiss pilot trial of early sevoflurane in severe COVID-19 lung injury randomised 60 patients across four hospitals to 48 hours of sevoflurane or intravenous sedation. Death or persistent organ dysfunction at day 28 occurred in 11/28 and 13/32 patients,

respectively. The trial was small and did not establish equivalence, but it did not demonstrate the hoped-for organ protection. Acute kidney injury was reported more often with sevoflurane, providing a cautionary finding before SESAR. Differences in case mix, duration and comparator prevent direct numerical pooling within a narrative appraisal.⁶

ROSE informs the accompanying sedation strategy rather than the choice of volatile agent. It randomised 1,006 patients with moderate to severe ARDS to early cisatracurium with deep sedation or usual care with lighter sedation targets, using high PEEP in both groups. Ninety-day in-hospital mortality was 42.5% versus 42.8%, and routine early paralysis did not improve survival. ROSE therefore cautions against treating deep sedation and paralysis as universally necessary background care. It does not invalidate SESAR's internal comparison, but helps define where its strategy differs from another evidence-based approach to similar hypoxaemia.⁷

Taken together, the evidence illustrates the separation between delivering sedation, improving a surrogate and improving patients' outcomes. The early ARDS pilot justified a larger study; it did not predict its result reliably. Sedaconda demonstrated sedative performance for a different volatile drug and treatment duration. Smaller hypoxaemia and COVID-19 studies did not provide persuasive evidence of lung or organ protection. SESAR supplies the largest direct randomised comparison of the proposed sevoflurane strategy with propofol in this ARDS population. By the end of 2025, its adverse primary result and concerning survival finding outweigh the earlier physiological promise for routine use of this strategy.

Should sevoflurane replace propofol for sedation in early moderate to severe ARDS?

No. The tested sevoflurane strategy reduced ventilator-free days and raised substantial concern about mortality and renal harm.

References

1. Jabaudon M, Boucher P, Imhoff E, Chabanne R, Faure JS, Roszyk L, et al. Sevoflurane for Sedation in Acute Respiratory Distress Syndrome. A Randomized Controlled Pilot Study. *Am J Respir Crit Care Med*. 2017;195(6):792-800.

2. Meiser A, Volk T, Wallenborn J, Guenther U, Becher T, Bracht H, et al. Inhaled isoflurane via the anaesthetic conserving device versus propofol for sedation of invasively ventilated patients in intensive care units in Germany and Slovenia: an open-label, phase 3, randomised controlled, non-inferiority trial. *Lancet Respir Med*. 2021;9(11):1231-1240.
3. Jabaudon M, Quenot JP, Badie J, Audard J, Jaber S, Rieu B, et al. Inhaled Sedation in Acute Respiratory Distress Syndrome: The SESAR Randomized Clinical Trial. *JAMA*. 2025;333(18):1608-1617.
4. Jabaudon M, Quenot JP, Badie J, Audard J, Jaber S, Rieu B, et al; SESAR Trial Investigators. Supplement 2: supplemental methods and results to Inhaled Sedation in Acute Respiratory Distress Syndrome: The SESAR Randomized Clinical Trial. *JAMA*. 2025.
5. Becher T, Meiser A, Guenther U, Bellgardt M, Wallenborn J, Kogelmann K, et al. Isoflurane vs. propofol for sedation in invasively ventilated patients with acute hypoxemic respiratory failure: an a priori hypothesis substudy of a randomized controlled trial. *Ann Intensive Care*. 2022;12(1):116.
6. Beck-Schimmer B, Schadde E, Pietsch U, Filipovic M, Dübendorfer-Dalbert S, Fodor P, et al. Early sevoflurane sedation in severe COVID-19-related lung injury patients. A pilot randomized controlled trial. *Ann Intensive Care*. 2024;14(1):41.
7. National Heart, Lung, and Blood Institute PETAL Clinical Trials Network; Moss M, Huang DT, Brower RG, Ferguson ND, Ginde AA, Gong MN, et al. Early Neuromuscular Blockade in the Acute Respiratory Distress Syndrome. *N Engl J Med*. 2019;380(21):1997-2008.

Possible Reporting Inconsistencies. The supplied principal report lists 700 patients randomised in its flow diagram but describes 687 as randomised in the abstract; 687 is the analysed population. Its Table 1 also gives 16 (34.0%) for propofol patients randomised on the day of ICU admission, an internally inconsistent count and percentage. The supplied editorial gives sevoflurane 90-day mortality as 53.2% and describes the COVID-19 interaction as non-significant; the principal report and supplement give 52.9% and a 90-day interaction P value of 0.01. This chapter uses the principal outcome table and supplementary subgroup analysis.

UK-ROX

Martin DS, Gould DW, Shahid T, Doidge JC, Cowden A, Sadique Z, et al. Conservative Oxygen Therapy in Mechanically Ventilated Critically Ill Adult Patients: The UK-ROX Randomized Clinical Trial. JAMA. 2025;334(5):398-408.

Introduction

Oxygen is essential treatment for hypoxaemia, but giving more than is needed may expose patients to pulmonary and systemic harm. Reactive oxygen species, vasoconstriction and absorption atelectasis provide plausible reasons to avoid unnecessary exposure. At the other end of the range, accepting lower oxygen saturation reduces the margin for measurement error, transient deterioration and inadequate tissue oxygen delivery. The clinical task is to find a useful balance, not simply to minimise a laboratory or monitor value.

Previous randomised trials had not established a broadly applicable oxygen target for mechanically ventilated adults. A small single-centre study suggested a survival advantage from avoiding hyperoxaemia, while larger subsequent studies found no clear benefit from lower oxygenation strategies. Those trials differed in populations, prescribed targets, achieved separation and control treatment. The label “conservative oxygen” therefore described several different interventions.^{1,2,3,4}

Usual practice had also changed. If contemporary clinicians already avoid marked hyperoxaemia, the additional benefit of imposing a much lower saturation target might be small. Detecting a modest mortality effect would require a substantially larger trial than most preceding critical care studies. Such an effect could nevertheless matter because oxygen is administered to so many patients.

UK-ROX addressed this question through a large registry-embedded comparison across United Kingdom intensive care units. It tested an explicit strategy of minimising supplemental oxygen while targeting a peripheral saturation of 90%, against contemporary clinician-directed oxygen therapy. Its scale offers unusually precise evidence for that policy decision, while its pragmatic design places limits on claims about a single biologically optimal saturation.⁵

Synopsis

UK-ROX was a pragmatic, individually randomised trial embedded within the Intensive Care National Audit and Research Centre Case Mix Programme. Ninety-seven adult general ICUs in England, Wales and Northern Ireland participated. Recruitment ran from 4 May 2021 to 27 November 2024, and 90-day follow-up was complete in February 2025. The National Institute for Health and Care Research funded the trial. Independent steering and data-monitoring committees provided oversight. The principal report was published online on 12 June 2025.^{5,6}

Adults were eligible within twelve hours of meeting the treatment criteria: invasive ventilation after an unplanned ICU admission, or ventilation initiated in the ICU, together with supplemental oxygen. Exclusions included extracorporeal membrane oxygenation, recent previous trial enrolment and a clinician judgement that either strategy was indicated or contraindicated. Research without prior consent allowed prompt inclusion, followed by consultation with representatives or consent from patients who regained capacity. Approval to obtain the primary mortality endpoint without subsequent linkage consent helped preserve follow-up, except when participants requested complete data removal.

A central telephone or web system concealed allocation to the two groups in a 1:1 ratio. Variable blocks were stratified by site and a hierarchical diagnostic classification: hypoxic-ischaemic encephalopathy, sepsis, other acute brain injury, or none of these. Clinical treatment was unblinded.

The conservative group received the lowest inspired oxygen concentration needed to target SpO₂ 90%, with an acceptable range of 88–92%. Monitor alarms were set below 88% and above 92%; the upper alarm could be disabled once the patient was breathing an inspired oxygen fraction of 0.21. Oxygen was reduced when saturation exceeded the range and increased when it fell below it. Exceptions were allowed for major disagreement between pulse oximetry and arterial measurements, acute life-threatening circumstances or a new contraindication. Other therapies could be altered according to clinical judgement.

Usual-care patients received oxygen at the treating clinician's discretion, without a mandated minimum inspired concentration or upper saturation alarm. This was not a protocol requiring sustained hyperoxaemia. The assigned policy continued until ICU discharge or day 90, with resumption advised after ICU readmission during that period.

All-cause mortality at 90 days was the primary outcome. Secondary endpoints included ICU and hospital mortality, mortality at 60 days and one year, time to discharge and days

alive without respiratory, cardiovascular or kidney support through day 30. Most information came from routine clinical datasets and national death registrations. Detailed hourly oxygen data were collected for up to ten days in an enhanced sample: initially the first ten patients at each site, followed by a random sample of subsequent enrolments, with extension where adherence required further monitoring.

The planned sample of 16,500 assumed 37% control mortality and sought an absolute reduction of 2.5 percentage points with 90% power, allowing 6% loss to follow-up. Two interim analyses used a Haybittle–Peto threshold of $P < 0.001$ for benefit or harm. The primary model adjusted for site, diagnostic group, age, baseline saturation and oxygenation ratio, COVID-19 status and calendar date. Marginal standardisation generated adjusted risk differences and relative risks. Missing data were addressed through imputation. Secondary and subgroup testing was not adjusted for multiplicity.

Of 52,747 screened patients, 16,500 were randomised: 8,258 conservative and 8,242 usual care. Sixty-six requested removal of all data. The primary analysis included 16,434 patients, with observed mortality available for 16,394 and imputation for forty whose records could not be linked. Median age was 60 years. Approximately one-third were classified as having sepsis, 9.2% hypoxic-ischaemic encephalopathy and 2.2% another acute brain injury. Median time from invasive ventilation in ICU to randomisation was five hours. Baseline inspired oxygen concentration was 0.45 in both groups.

Enhanced oxygen data were available for 2,489 patients. The mean of individual median inspired oxygen fractions was 0.31 with conservative therapy and 0.35 with usual care. Cumulative exposure above room air was 20.3 versus 28.7 equivalent hours of 100% oxygen, a 29.3% reduction. This dose metric combines concentration and duration; it does not mean that patients literally received those durations at an inspired concentration of 1.0.

The mean of individual median SpO_2 values was 93.3% versus 95.1%, and the corresponding PaO_2 values 71.5 versus 79.5 mm Hg. Conservative patients spent more time in the 88–92% range and more time on room air above that range. Mean time below 88% was 3.2 versus 2.3 hours. Among monitored conservative patients, 42.1% had at least one non-adherent period, accounting for 10.6% of ICU time. Staffing, competing priorities, clinical decisions and undocumented reasons contributed. Separation persisted across the trial's recruitment period.

By day 90, 2,908/8,211 conservative patients and 2,858/8,183 usual-care patients had died, 35.4% versus 34.9%. The adjusted risk difference was 0.7 percentage points (95% CI –0.7 to 2.0; $P=0.28$), compared with an unadjusted difference of 0.5 percentage points

(95% CI –1.0 to 2.0). There was no significant difference in organ-support-free days, with an adjusted proportional odds ratio of 1.01 (95% CI 0.96–1.07). Other mortality time points and ICU or hospital discharge outcomes were also similar.

Serious adverse events were reported in 58 conservative patients and 29 usual-care patients, 0.7% versus 0.4%. Prespecified interaction tests did not identify a convincing difference by diagnostic group, COVID-19 status or ethnic group. Post hoc severity analyses were similarly inconclusive. One-year follow-up was available only for those who had reached that interval before the final linkage date; it should not be confused with complete one-year observation of every randomised participant.^{5,6}

Critique

UK-ROX provides strong evidence against a clinically substantial mortality benefit from routinely lowering oxygen exposure to target SpO₂ 90% in this broad ICU population. Its adjusted interval allows, at most, a 0.7-percentage-point mortality reduction while also allowing a two-percentage-point increase. The result is therefore considerably more informative than the statement that P exceeded 0.05. It does not prove equivalence, but it excludes the magnitude of benefit for which the trial was designed.

The comparator is central. Usual care achieved a mean of patient median saturations around 95% and PaO₂ around 80 mm Hg. Those values do not describe sustained severe hyperoxaemia. The study tests whether lowering exposure further improves outcomes compared with already moderate contemporary practice. It cannot be used to justify unrestricted oxygen or to conclude that very high inspired concentrations and arterial tensions are harmless. Earlier trials with more liberal control strategies answer a partly different question.

The intervention achieved a meaningful reduction in oxygen exposure despite a modest saturation separation. Saturation lies on a nonlinear dissociation curve, and a patient can remain above 92% while breathing room air. The conservative policy did not require artificial reduction of saturation once supplemental oxygen had been removed. An achieved value above the target therefore does not automatically indicate protocol failure. The combination of lower inspired concentration, less cumulative exposure and lower arterial oxygen tension confirms that the randomised strategies produced different treatment.

Non-adherence nevertheless limits the distinction between policy effectiveness and biological efficacy under perfect delivery. More than two-fifths of monitored conservative patients had a deviation, although the proportion of time affected was

much smaller. A strictly enforced target might produce a larger physiological separation, but could also cause more hypoxaemia or require more clinical intervention. The present result is best understood as the effect of implementing the policy with realistic safety exceptions and bedside constraints. It does not establish the effect of an idealised intervention that clinicians may be unwilling or unable to maintain.

Detailed exposure was sampled rather than collected universally. That was a deliberate efficiency choice enabling the large trial, not evidence that the investigators failed to measure treatment altogether. The enhanced sample included both early implementation patients and randomly selected later patients. Similar separation across those groups and over calendar time supports the interpretation that the policy changed exposure more broadly. Still, measurements in approximately 15% cannot confirm each patient's oxygen dose or rule out all centre-specific differences in the remainder.

The mortality endpoint and near-complete ascertainment are major strengths. National linkage reduces dependence on participants returning for follow-up, and the approved consent framework preserved outcome collection for almost everyone who did not request complete withdrawal. Only forty primary outcomes required imputation after data-removal exclusions. This makes a large effect hidden by missing mortality data unlikely. Secondary outcomes had different linkage-consent requirements and should not automatically inherit the same completeness.

The large sample does not remove selection at enrolment. More than ten thousand screened patients were excluded because a strategy was considered indicated or contraindicated, and many otherwise eligible patients were missed or not randomised for clinical reasons. These decisions preserve appropriate care but limit applicability where clinicians already have strong reasons to choose a particular target. The trial is especially relevant to the substantial group in whom equipoise existed, rather than every mechanically ventilated patient regardless of diagnosis or circumstances.

Pulse oximetry is a practical but imperfect treatment guide. Measurement error, poor peripheral perfusion and discordance with arterial saturation become more consequential near a low target. The protocol allowed arterial values to inform departures from the assigned range, an important safety feature. A neutral interaction by ethnic group does not prove identical measurement performance in all patients or validate low saturation targets for every subgroup, particularly where numbers are limited. The trial tests the whole clinical strategy, including these measurement limitations.

Subgroups should be interpreted in relation to their interaction tests and precision. Sepsis and post-cardiac-arrest encephalopathy are biologically plausible populations of interest, but plausibility alone does not identify a group that benefits from conservative therapy. The small other-brain-injury group is particularly insufficient for a firm diagnosis-specific recommendation. Future individual-treatment-effect models would require external validation and prospective evaluation before determining bedside oxygen targets.

The numerical excess of reported serious adverse events deserves acknowledgement, alongside the somewhat greater time below 88%. It is not a clean estimate of all oxygen-related harm: reporting in an unblinded pragmatic trial can be influenced by attribution, and event counts were small relative to the sample. Nor does the neutral mortality result establish absence of nonfatal injury. The trial supports neither a definitive overall harm claim nor a blanket statement that lowering the target is equally safe.

The clinical consequence is straightforward but bounded. In patients resembling those enrolled, there is no demonstrated survival reason to impose an SpO₂ target of 90% rather than contemporary clinician-directed practice. Oxygen should still be titrated and unnecessary excess avoided. UK-ROX narrows an important uncertainty about a routine policy; it does not locate one universally optimal saturation or replace disease-specific judgement where the trial population and comparison do not apply.

Body of Evidence

Oxygen-ICU initially generated enthusiasm for conservative treatment. This single-centre trial analysed 434 patients after stopping before its planned recruitment target. Conservative treatment aimed for PaO₂ 70–100 mm Hg or SpO₂ 94–98%, compared with a more liberal strategy permitting PaO₂ up to 150 mm Hg or saturation 97–100%. ICU mortality was 11.6% versus 20.2%, favouring conservative treatment. The size of that effect, early termination and single-centre design required replication. Its conservative saturation range also overlaps substantially with the usual-care values subsequently observed in UK-ROX.¹

ICU-ROX randomised 1,000 mechanically ventilated patients to conservative or usual oxygen. Both strategies used a default lower saturation limit of 90%; the conservative arm restricted unnecessary oxygen and used an upper alarm at 97%. Median ventilator-free days were 21.3 versus 22.1, with no significant difference, and 180-day mortality was 35.7% versus 34.5%. This showed that reducing exposure could be

achieved without demonstrating improved recovery, but the mortality interval remained much wider than UK-ROX's.²

HOT-ICU enrolled 2,928 patients with acute hypoxaemic respiratory failure, including patients on open oxygen systems and those receiving higher inspired concentrations in closed systems. It compared PaO₂ targets of 60 and 90 mm Hg. Ninety-day mortality was 42.9% versus 42.4%, adjusted risk ratio 1.02 (95% CI 0.94–1.11). Its arterial-pressure target and selected respiratory-failure population differ from UK-ROX, but both trials failed to demonstrate improved survival from lower oxygenation.³

PILOT used a cluster-randomised crossover design at one academic centre to compare saturation targets of 90%, 94% and 98%, with respective ranges of 88–92%, 92–96% and 96–100%. Among 2,541 patients, median ventilator-free days were 20, 21 and 21, and in-hospital mortality by day 28 was 34.8%, 34.0% and 33.2%. There was no clear difference in the primary outcome. Its explicit three-target comparison helps contextualise the range of clinical strategies, while UK-ROX supplies much greater precision for a low-target policy against usual care.⁴

LOCO₂ provides a caution against assuming that lower oxygenation is inherently safer in severe lung injury. In patients with acute respiratory distress syndrome, it compared PaO₂ 55–70 mm Hg and SpO₂ 88–92% with higher targets. Recruitment stopped after 205 patients because of safety concerns and low likelihood of demonstrating the primary benefit. The 28-day mortality comparison was inconclusive, but 90-day mortality was higher in the conservative group and five mesenteric ischaemic events occurred there. The small, prematurely stopped study does not precisely quantify harm, but its findings argue against unrestricted extrapolation of low targets.⁷

Together, these trials show why oxygen strategy must be defined by actual targets, exposure and comparator practice. UK-ROX adds a very large, broadly applicable randomised estimate and substantially excludes the prospect of a large general survival gain from targeting 90% in contemporary UK intensive care. The accompanying editorial also highlights the research achievement: routine clinical infrastructure can support trials large enough to address modest effects of common therapies. That methodological success strengthens the answer to the tested question without expanding it beyond the population and strategies actually studied.⁸

Should ventilated ICU adults routinely have oxygen reduced to target SpO₂ 90% to improve survival?

No. UK-ROX found no survival benefit over contemporary usual care; this does not justify unnecessary hyperoxaemia or settle targets for every specific condition.

References

1. Girardis M, Busani S, Damiani E, Donati A, Rinaldi L, Marudi A, et al. Effect of Conservative vs Conventional Oxygen Therapy on Mortality Among Patients in an Intensive Care Unit: The Oxygen-ICU Randomized Clinical Trial. *JAMA*. 2016;316(15):1583-1589.
2. Mackle D, Bellomo R, Bailey M, Beasley R, Deane A, Eastwood G, et al. Conservative Oxygen Therapy during Mechanical Ventilation in the ICU. *N Engl J Med*. 2020;382(11):989-998.
3. Schjørring OL, Klitgaard TL, Perner A, Wetterslev J, Lange T, Siegemund M, et al. Lower or Higher Oxygenation Targets for Acute Hypoxemic Respiratory Failure. *N Engl J Med*. 2021;384(14):1301-1311.
4. Semler MW, Casey JD, Lloyd BD, Hastings PG, Hays MA, Stollings JL, et al. Oxygen-Saturation Targets for Critically Ill Adults Receiving Mechanical Ventilation. *N Engl J Med*. 2022;387(19):1759-1769.
5. Martin DS, Gould DW, Shahid T, Doidge JC, Cowden A, Sadique Z, et al. Conservative Oxygen Therapy in Mechanically Ventilated Critically Ill Adult Patients: The UK-ROX Randomized Clinical Trial. *JAMA*. 2025;334(5):398-408.
6. Martin DS, Gould DW, Shahid T, Doidge JC, Cowden A, Sadique Z, et al. Supplement 2 to Conservative Oxygen Therapy in Mechanically Ventilated Critically Ill Adult Patients: The UK-ROX Randomized Clinical Trial. *JAMA*. 2025.
7. Barrot L, Asfar P, Mauny F, Winiszewski H, Montini F, Badie J, et al. Liberal or Conservative Oxygen Therapy for Acute Respiratory Distress Syndrome. *N Engl J Med*. 2020;382(11):999-1008.
8. Seitz KP, Casey JD, Semler MW. Patient, Treatment, Outcome—Large Simple Trials of Common Therapies. *JAMA*. 2025;334(5):395-397.

PROMIZING

Bosma KJ, Burns KEA, Martin CM, Skrobik Y, Mancebo Cortés J, Mulligan S, et al. Proportional-Assist Ventilation for Minimizing the Duration of Mechanical Ventilation. N Engl J Med. 2025;393(11):1088-1103.

Introduction

Liberation from invasive ventilation requires both recovery from the underlying illness and sufficient respiratory capacity to sustain breathing. Excessive assistance can suppress respiratory muscle activity, whereas inadequate assistance can produce distress and potentially injurious effort. Pressure-support ventilation is widely used during recovery, but the delivered pressure does not automatically adapt breath by breath to the patient's changing effort. A comfortable respiratory rate and tidal volume do not necessarily guarantee an appropriate muscular workload.

Proportional-assist ventilation with load-adjustable gain factors, PAV+, attempts to address this limitation. The ventilator estimates respiratory mechanics and supplies a selected proportion of the pressure needed to generate each breath. Assistance consequently varies with effort rather than delivering the same pressure increment for every inspiration. Clinicians can use the displayed estimate of patient effort to adjust assistance, aiming to avoid both excessive unloading and excessive work. This is an attractive physiological strategy, but improvements in synchrony or comfort need not shorten the time to independent breathing.¹

Earlier small studies suggested that PAV+ could reduce dyssynchrony, interruptions of spontaneous support and bedside adjustments. Evidence for accelerated successful liberation was much less secure. A mode that performs well under expert physiological observation may offer little additional benefit over carefully titrated pressure support and a reliable programme of spontaneous breathing trials.^{2,3}

PROMIZING tested whether PAV+ shortened ventilation in patients able to breathe with partial assistance but not yet ready for liberation. Both groups received explicit support-adjustment algorithms and the same structured liberation pathway. This makes the trial especially informative about the added value of the mode when the surrounding process is already organised.¹

Synopsis

PROMIZING was an investigator-initiated, open-label, randomised superiority trial at 23 hospitals in Argentina, Canada, France, Greece, Italy, Saudi Arabia and Spain. Recruitment ran from September 2016 to June 2023, with interruption during the COVID-19 pandemic. Support came from the Canadian Institutes of Health Research, including an industry-partnered grant with Covidien/Medtronic, and the University of Western Ontario. An independent committee monitored safety and adherence. Sites completed a run-in training phase using both modes before randomising patients.^{1,4}

Adults ventilated for at least 24 hours were eligible if they could tolerate pressure-support ventilation but were not ready for liberation. Severe chronic respiratory disease, major neurological disorders and anticipated withdrawal of life support were important exclusions. Screening involved several stages. Patients had to tolerate at least thirty minutes of pressure support and then demonstrate that they were not ready for spontaneous breathing: failure to meet weaning criteria, a rapid shallow-breathing index above 100 during a brief unsupported assessment, or failure of a thirty-minute spontaneous breathing trial. Patients already on pressure support for more than 24 hours, passing the breathing trial or expected to be extubated within 24 hours were excluded.

Central computer-generated randomisation used concealed allocation, variable blocks and centre stratification. Assignment was equal between PAV+ and pressure support. Bedside clinicians and research assistants knew the mode, but the trial statistician was masked. Consent arrangements included prospective or deferred consent according to the applicable circumstances.

PAV+ was delivered using Puritan Bennett 840 or 980 ventilators. Gain was adjusted to keep estimated peak patient-generated respiratory muscle pressure between 5 and 10 cm H₂O. Pressure support could be delivered on any ventilator and was adjusted to maintain a respiratory rate of 12–35 breaths/min and tidal volume of 5–10 mL/kg predicted body weight. Airway occlusion pressure at 0.1 seconds was recorded but did not determine adjustments. Algorithms provided responses to hypoventilation, hyperventilation and distress, including temporary return to assist-control ventilation when necessary.

Both arms underwent the same daily assessment for a spontaneous breathing trial. Trials used a T-piece or equivalent unsupported method with an inspired oxygen fraction of 0.40. Once a patient passed and met extubation criteria, extubation was intended within two hours. Tracheostomy remained a clinical decision. Guidance addressed sedation, nutrition and mobilisation, but non-respiratory treatment was not rigidly protocolised.

Daily records documented support, liberation decisions and whether staff found the allocated strategy easy to use.

The primary endpoint was time from randomisation to successful liberation, defined as extubation or final disconnection followed by seven days alive without invasive ventilation. Follow-up continued until liberation, live ICU discharge, death or day 90. Death was treated as a competing event. Secondary outcomes included mortality, discharge, ventilator-free days, weaning progress, reintubation, tracheostomy, return to controlled ventilation, sedative exposure and an ordinal combination of living situation and ventilation status at day 90.

The original primary endpoint of ventilator-free days at day 21 was changed early after blinded aggregate review indicated that it inadequately represented patients still ventilated at that point. Aggregate data from the first 120 patients also informed sample-size reassessment. A minimum of 558 randomised patients was considered feasible and would provide 80% power to detect a reduction in median ventilation duration from 7.70 to 5.92 days, corresponding to a hazard ratio of 1.30. There were no formal interim efficacy analyses. The principal comparison used cumulative incidence curves and Gray's test; a multistate model estimated transitions including liberation and death. Secondary intervals were not multiplicity-adjusted.

Of 10,557 screened patients, 722 entered the staged enrolment process and 575 were randomised. Two assigned PAV+ were excluded before starting treatment, leaving 287 PAV+ and 286 pressure-support patients in the analysis. Three PAV+ patients withdrew consent for continued data collection and were censored at last contact. The most common reason for enrolled patients not reaching randomisation was already passing a spontaneous breathing trial or being extubated. Median ventilation duration before randomisation was approximately five days; almost half had a respiratory cause for ICU admission and about half were receiving vasopressors.

Initial pressure support averaged 10.8 cm H₂O, while initial PAV+ gain averaged 65%. Respiratory rates and tidal volumes were similar. In the PAV+ arm, estimated respiratory muscle pressure was in the intended range on 80.3% of measured trial days, above it on 12.8% and below it on 7.0%. Airway occlusion pressure had a similar distribution in both groups. Non-protocolised modes accounted for 2.5% of trial days. More PAV+ patients discontinued the assigned intervention before completion, 27 versus nine, for reasons including transfer, mode violations and withdrawal.

Median time to successful liberation was 7.3 days (95% CI 6.2–9.7) with PAV+ and 6.8 days (95% CI 5.4–8.8) with pressure support; Gray's test gave P=0.58. The multistate

hazard ratio for transition from ventilation to liberation in ICU was 0.96 (95% CI 0.80–1.15). The reported difference in median times was 0.5 days (95% CI –3.0 to 3.5). Sensitivity analyses using 48 hours without renewed ventilation and per-protocol analyses were consistent with the main result.

Day-90 mortality was 29.6% versus 26.6%, a difference of 3.0 percentage points (95% CI –4.3 to 10.4). Reintubation or reconnection occurred in 22.2% versus 23.3% of the relevant patients, and post-randomisation tracheostomy in 20.9% versus 19.2%. Ventilator-free days and discharge outcomes did not establish benefit. The day-28 change in midazolam-equivalent dose relative to baseline differed by –1.55 mg/kg between groups, but its 95% confidence interval extended from –4.92 to 1.81. Serious adverse events affected 10.8% versus 9.8% (P=0.79). Staff reported greater ease with pressure support, and return to assist–control occurred more often with PAV+.^{1,4}

Critique

PROMIZING found no evidence that PAV+ accelerates sustained liberation compared with carefully managed pressure support. The primary point estimate was close to neutral, and related clinical endpoints did not reveal a consistent advantage. This is a substantial test of the proposed clinical benefit, even though it cannot exclude smaller effects or benefit from different implementation in selected patients.

The control strategy is one of the trial's most important strengths. Pressure support was adjusted using explicit targets and linked to daily breathing trials and prompt extubation. A benefit arising merely from more attentive weaning in the intervention group was therefore less likely. The same feature limits comparison with poorly organised usual care. PROMIZING does not show that the surrounding protocols are ineffective, nor that either mode would perform identically where spontaneous breathing trials are delayed or support is rarely reassessed.

The selection process demonstrates this distinction. Many enrolled patients proved ready for spontaneous breathing before they could be randomised. Systematic assessment itself identified opportunities to stop ventilation that had not yet been recognised. Those patients could not benefit from prolonged exposure to either weaning mode. Excluding them appropriately concentrated the comparison on patients still needing assistance, but also emphasises that improving recognition of readiness may be more immediately useful than changing ventilator technology.

The hypothesis that PAV+ would produce more appropriate effort was incompletely separated from the comparator in practice. Estimated muscle pressure was usually

within the intended range under PAV+, but airway occlusion pressure, a marker of respiratory drive, was distributed similarly under both strategies. This is compatible with pressure support being well titrated. It does not prove identical diaphragm loading: occlusion pressure is not a complete measure of effort, and patient muscle pressure was not continuously measured by the same method in both groups. A neutral outcome could reflect an effective comparator, insufficient physiological differentiation, or absence of an outcome benefit from the proposed mechanism.

Timing also matters. Although the intervention was introduced early in the transition to partial support, patients had already received ventilation for approximately five days. Diaphragm dysfunction or other consequences of preceding critical illness could already have developed. The trial does not test PAV+ from the first hours of ventilation, and it would be inappropriate to interpret its result as disproving every strategy aimed at preserving early respiratory muscle activity. Equally, an untested earlier strategy cannot be used to claim that the neutral trial supports routine adoption.

The endpoint is clinically stronger than time to first extubation alone. Requiring seven days alive without renewed invasive ventilation avoids counting a brief unsuccessful attempt as recovery. Accounting for death as a competing event also recognises that death prevents liberation. The multistate hazard ratio describes a particular transition within that framework; it is not a percentage reduction in total ventilation days or a measure of muscle recovery. The median difference and its broad interval supply a complementary view of the uncertainty.

The primary-outcome change deserves transparent reporting. The main paper states that it followed blinded aggregate review early in recruitment, rather than an examination of treatment differences. That reduces concern about choosing an endpoint because it favoured one arm. It does not remove the need to document the change, and a separate protocol audit would be needed to independently reconstruct every amendment. The resulting endpoint is reasonable, but the trial should be appraised as the amended comparison actually conducted.

Its analysis was labelled intention-to-treat while excluding two randomised patients who never started treatment. That is a modified population, although the small number makes a major distortion less likely. More consequential for supportive analyses is the imbalance in intervention discontinuation. A per-protocol comparison conditions on events after randomisation, including transfer and tolerance, and cannot automatically be treated as a less biased estimate of efficacy. The primary allocation-based analysis remains the main guide.

PAV+ requires equipment, training and familiarity. Centres completed a run-in, yet staff found pressure support easier and switched from PAV+ to controlled ventilation more often. Limited familiarity could reduce effectiveness, but implementation difficulty is also part of the real clinical strategy. The trial cannot distinguish completely between an intrinsic limitation of the mode and a limitation of delivery. Its results do not justify assuming that another device platform, algorithm or training programme will reproduce better outcomes without direct evaluation.

Sedation findings should remain hypothesis-generating. The abstract highlights a larger reduction from baseline with PAV+, but the between-group confidence interval includes no difference and substantial uncertainty. Doses at late time points concern patients still available for measurement, and sedation was not the randomised intervention. Delirium assessments were incomplete and unavailable in deep sedation. Neither observation establishes that PAV+ reduces delirium or improves neurocognitive recovery, especially when the major clinical endpoints remain neutral.

Generalisation is strongest to adults who can tolerate partial support but fail a structured readiness assessment. Severe neurological disease, severe chronic respiratory illness and anticipated palliation were excluded. The results do not directly address every prolonged-weaning population, patients with minimal respiratory drive or those requiring fully controlled ventilation. They also do not establish equivalence of all proportional modes: PAV+ estimates mechanics and effort differently from neurally adjusted ventilatory assist.

The practical conclusion is therefore to value the liberation process over an assumed technological advantage. PAV+ may remain useful where clinicians find a specific physiological or comfort benefit, but PROMIZING provides no demonstrated reduction in ventilation duration to support routinely replacing well-managed pressure support for that purpose.

Body of Evidence

Xirouchaki and colleagues randomised 208 patients capable of assisted breathing to PAV+ or pressure support for up to 48 hours. Failure requiring controlled ventilation occurred in 11.1% versus 22.0%, and major dyssynchrony in 5.6% versus 29.0%. The proportion meeting criteria for unsupported breathing did not differ. This distinction is central: improved tolerance and synchrony established a physiological rationale, but did not themselves demonstrate faster liberation. The short observation window and experienced single-centre setting also differ from PROMIZING's longer multicentre clinical comparison.²

A related analysis of that same cohort found fewer ventilator-setting and sedative adjustments with PAV+. It is not independent replication of the original trial. Reduced adjustment burden could be valuable, but depends on local expertise and cannot be assumed from a device's operating principle. PROMIZING's finding that clinicians more often preferred pressure support illustrates how a technology's practical effect can differ between a specialist research environment and broader deployment.^{1,5}

Botha and colleagues conducted a fifty-patient pilot comparison during weaning. Time to successful liberation was numerically shorter with PAV+ but not significantly different, and four patients crossed from PAV+ to pressure support. The small trial was useful for feasibility and hypothesis formation; its sparse mortality observations cannot establish a survival effect. PROMIZING provides a much larger and more precise direct test of the liberation question.³

Neurally adjusted ventilatory assist, NAVA, is a related but distinct approach that links assistance to diaphragm electrical activity. In a 306-patient multicentre randomised trial, median ventilator-free days were 22 with NAVA and eighteen with conventional lung-protective ventilation, a reported difference of four days (95% CI 0–8; P=0.016). Hospital mortality did not differ clearly. The control group could receive several conventional modes at clinician discretion, whereas PROMIZING compared two explicitly managed partial-support strategies. The positive NAVA result therefore cannot be transferred directly to PAV+ or interpreted as a contradiction without considering the different technology and comparator.⁶

The cumulative evidence supports separating three questions: whether a mode improves interaction between patient and ventilator, whether clinicians can use it consistently, and whether it improves sustained recovery. PAV+ has supportive evidence for selected physiological outcomes, but PROMIZING did not translate that rationale into faster liberation under a structured weaning programme. It narrows the case for routine mode replacement while leaving specific physiological indications and alternative implementation strategies open to further study.

Should PAV+ replace protocolised pressure support to shorten invasive ventilation?

Not routinely. PROMIZING found no faster sustained liberation when both modes were carefully adjusted within the same structured weaning programme.

References

1. Bosma KJ, Burns KEA, Martin CM, Skrobik Y, Mancebo Cortés J, Mulligan S, et al. Proportional-Assist Ventilation for Minimizing the Duration of Mechanical Ventilation. *N Engl J Med*. 2025;393(11):1088-1103.
2. Xirouchaki N, Kondili E, Vaporidi K, Xirouchakis G, Klimathianaki M, Gavriilidis G, et al. Proportional assist ventilation with load-adjustable gain factors in critically ill patients: comparison with pressure support. *Intensive Care Med*. 2008;34(11):2026-2034.
3. Botha J, Green C, Carney I, Haji K, Gupta S, Tiruvoipati R. Proportional assist ventilation versus pressure support ventilation in weaning ventilation: a pilot randomised controlled trial. *Crit Care Resusc*. 2018;20(1):33-40.
4. Bosma KJ, Burns KEA, Martin CM, Skrobik Y, Mancebo Cortés J, Mulligan S, et al. Supplementary appendix to Proportional-Assist Ventilation for Minimizing the Duration of Mechanical Ventilation. *N Engl J Med*. 2025.
5. Xirouchaki N, Kondili E, Klimathianaki M, Georgopoulos D. Is proportional-assist ventilation with load-adjustable gain factors a user-friendly mode? *Intensive Care Med*. 2009;35(9):1599-1603.
6. Kacmarek RM, Villar J, Parrilla D, Alba F, Solano R, Liu S, et al. Neurally adjusted ventilatory assist in acute respiratory failure: a randomized controlled trial. *Intensive Care Med*. 2020;46(12):2327-2337.

ACTIVE

Sinnige JS, Buiteman-Kruizinga LA, Horn J, Paulus F, Schultz MJ, Serpa Neto A. Effect of Automated Closed-Loop Ventilation vs Protocolized Conventional Ventilation on Ventilator-Free Days in Critically Ill Adults: A Randomized Clinical Trial. JAMA. 2026;335(10):874-884. Published online December 8, 2025.

Introduction

Mechanical ventilation requires repeated decisions about tidal volume, respiratory rate, pressure, oxygen and readiness to resume unsupported breathing. A setting that is appropriate now may become excessive or inadequate as respiratory mechanics, metabolic demand and consciousness change. Even where clinicians know the desired targets, staffing and competing priorities can delay adjustment. Automation offers the possibility of more consistent treatment through continuous measurement and rapid response.

INTELLiVENT adaptive support ventilation combines automated control of ventilation and oxygenation. It adjusts the breathing pattern to estimated respiratory mechanics and carbon dioxide targets, while changing inspired oxygen and positive end-expiratory pressure in response to oxygen saturation. The underlying idea is to maintain acceptable gas exchange with less unnecessary ventilatory intensity and to facilitate transition towards spontaneous breathing. Clinicians still choose relevant targets, evaluate sensor reliability and remain responsible for the patient.¹

Physiological optimisation and clinical recovery are different outcomes. A system can keep more breaths within predefined limits without changing the underlying disease, consciousness or other barriers to extubation. Conversely, delivering similar clinical outcomes with less staff effort could be valuable even without superiority, provided workload savings and safety are actually established. Earlier small studies had suggested improved physiological control, but did not settle either question in a broad critically ill population.²

ACTIVE tested whether initiating closed-loop ventilation very early increased ventilator-free days compared with protocolised conventional ventilation. Both groups received structured sedation and liberation care. The resulting comparison assesses the additional clinical benefit of a specific automation platform within experienced units, rather than automation against inconsistent or infrequently reviewed care.

Synopsis

ACTiVE was an investigator-initiated, open-label, randomised superiority trial at seven ICUs in the Netherlands and Switzerland. Recruitment occurred between October 2020 and June 2025. ZonMw funded the work, with Amsterdam University Medical Centers coordinating and sponsoring the trial. The funder had no role in design, conduct, analysis or publication decisions. An independent data and safety monitoring board reviewed conduct and safety. The principal report appeared online on 8 December 2025.^{1,3}

Adults were eligible within one hour of starting invasive ventilation in the ICU if ventilation was expected to continue for at least 24 hours. Patients ventilated for more than six hours before ICU admission were excluded. Other important exclusions were extracorporeal support, unreliable pulse oximetry, body mass index above 40, restrictive lung disease, recent major lung resection and neuromuscular disease expected to prolong ventilation. An appropriate ventilator had to be available. Patients were randomised under deferred-consent arrangements, but applicable rules required removal of their data if subsequent consent could not be obtained.

Central web randomisation allocated patients 1:1 with concealed, variable-sized blocks stratified by centre. Clinical staff and investigators knew the strategy; the statistician was masked. The intervention was initiated promptly after allocation and continued until successful extubation, death or day 28.

In the automated arm, INTELLiVENT-ASV selected a combination of tidal volume and respiratory rate intended to minimise breathing work and mechanical power for the required minute ventilation. End-tidal carbon dioxide guided ventilation, while pulse oximetry guided oxygen concentration and positive end-expiratory pressure. Optional presets adjusted targets for acute respiratory distress syndrome, chronic hypercapnia or brain injury. Staff could fine-tune targets. The QuickWean function was recommended, and automated spontaneous breathing trials were optional. If signals failed or targets could not be maintained, alarms prompted review and the system reverted to default or manual settings.

The conventional arm used volume control, pressure control or pressure support. Other automated modes were prohibited. Both groups followed lung-protective principles, including low tidal volumes, pressure limitation and appropriate oxygenation and end-expiratory pressure. Analgesia and avoidance of unnecessary deep sedation were emphasised. Conventional patients were assessed at least three times daily for transition to pressure support; both groups had extubation readiness evaluated at least three times daily. Clinicians made the final extubation decision using common physiological and clinical criteria.

The primary outcome was ventilator-free days through day 28. Patients dying before day 28 or ventilated beyond it received zero; otherwise days followed the last successful extubation, with at least 24 hours free of invasive ventilation required. Secondary outcomes included mortality, ventilation duration among survivors, lengths of stay, complications and ventilation quality. Quality classified breaths as optimal, acceptable or critical according to tidal volume, maximum airway pressure, end-tidal carbon dioxide and oxygen saturation; the worst component determined the overall category.

The planned analysed sample of 1,200 provided 80% power to detect a 1.5-day increase in ventilator-free days, assuming a mean of twenty days and standard deviation of nine, with 5% dropout. The primary analysis instead used the ordered distribution of ventilator-free days in a mixed-effects cumulative logistic model, with centre as a random effect. An odds ratio above one favoured automation. Missing outcomes were not imputed. Per-protocol analysis required the assigned strategy for at least 80% of ventilation time during the first five days. Secondary P values were adjusted with Holm–Bonferroni methods; quality-zone comparisons also used a specified multiplicity threshold.

Of 3,641 screened patients, 1,514 were randomised: 752 automated and 762 conventional. Data from 150 and 163, respectively, were removed because deferred consent was not obtained, leaving 602 and 599 in the modified intention-to-treat population. Primary outcome data were available for 601 and 595. Median age was 63 years, approximately 36% were women and median time from ventilation initiation to enrolment was 0.6 hours. About one-third were intubated for respiratory failure, one-quarter after cardiac arrest and one-quarter for neurological reasons. Baseline tidal volume was approximately 6.5 mL/kg predicted body weight and driving pressure approximately 13 cm H₂O.

Treatment separation was substantial. Median time under the assigned strategy was 96% in the automated group and 100% in controls. Ventilation changed promptly after randomisation, with lower minute ventilation and more conservative oxygenation in the automated arm. Sedation practice, fluid balance and organ-failure scores were broadly similar. Specialised disease presets were used in only a minority of the patients for whom sufficiently detailed device data were available.

Median ventilator-free days were 16.7 with automation and 16.3 with conventional ventilation. The primary proportional odds ratio was 0.91 (95% CI 0.77–1.06; P=0.23), indicating no improvement across the ordered outcome distribution. Mean values were 13.1 and 13.4 days, with a difference of –0.3 days (95% CI –1.7 to 1.0). The apparent numerical advantage in medians should therefore not be read as a favourable primary

effect. Adjusted and per-protocol analyses were consistent, and no convincing prespecified subgroup interaction was identified.

Day-28 mortality was 224/602 versus 218/597, 37.2% versus 36.5%; the hazard ratio was 1.04 (95% CI 0.86–1.25). Median ventilation duration among survivors was 3.3 versus 2.6 days, without a multiplicity-adjusted difference. Other clinical recovery outcomes were similar. In the small subset with analysable breath-level quality data, automation improved the ordered quality measure, with an odds ratio of 1.50 (95% CI 1.43–1.57), and reduced time classified as critical. Nominal reductions in rescue therapies and severe hypoxaemia did not remain statistically significant after adjustment for multiple testing.^{1,3}

Critique

ACTiVE does not demonstrate that early INTELLiVENT-ASV improves ventilator-free days over protocolised conventional ventilation. Its point estimate for the primary ordinal analysis favoured the control strategy, although the interval included no effect. Improved physiological classifications therefore should not be substituted for a clinical benefit that the trial did not show.

The headline medians are potentially misleading when viewed in isolation. A median of 16.7 rather than 16.3 days looks favourable, but the full distribution includes many deaths assigned zero and many short ventilation courses. The proportional odds model evaluates that ordered distribution rather than simply subtracting medians. Mean ventilator-free days and the primary odds ratio both point slightly towards conventional care. None establishes harm, but together they explain why the trial cannot be described as showing even a clear numerical improvement in its principal estimand.

The most important threat to internal validity is the loss of 313 randomised patients before analysis. Legal requirements explain why those data were removed; they do not make selection bias impossible. Consent availability can relate to survival, family contact, neurological status or illness trajectory. Because the trial was unblinded, it cannot be assumed solely from similar exclusion proportions that the retained groups preserve every benefit of randomisation. Baseline and outcome data from excluded patients were unavailable to test that assumption. This concern is distinct from the very small number of missing outcomes among the retained 1,201 patients.²

At the same time, the exclusions do not prove a biased result or establish its direction. The reported rates were reasonably balanced, and allocation was concealed at enrolment. The appropriate response is to reduce certainty about the complete

randomised population and acknowledge the unresolved selection mechanism, rather than claim that the observed neutral result is necessarily wrong. Future trials need consent and data-retention arrangements that protect participants while preserving a credible allocation-based analysis.

The comparator sets a demanding standard. Conventional ventilation was protocolised, lung-protective settings were emphasised and readiness for liberation was assessed at least three times daily. In experienced units, this may leave relatively little correctable delay for automation. The result is useful for deciding whether automation improves outcomes over well-delivered care. It does not establish whether the same system could improve consistency where staffing or expertise is more limited. That is a different comparison requiring direct evaluation, including the availability of people able to respond when the system fails.

Case mix also affects the opportunity for benefit. Roughly half the cohort was intubated following cardiac arrest or for neurological disease. Such patients may have adequate respiratory mechanics yet remain intubated because of consciousness, airway protection or neurological prognosis. Fine adjustment of ventilation may not change those barriers. Many survivors were liberated after only a few days, reducing exposure time during which a more adaptive strategy might matter. These observations plausibly limit the average treatment effect, but do not identify a proven benefiting respiratory-failure subgroup.

The small quality substudy is insufficient to explain the whole trial. The principal table contains analysable quality results for 142 patients, sampled non-systematically. A breath-level model can produce a very narrow confidence interval because it incorporates many observations, but thousands of breaths do not become thousands of independently randomised patients. The clinical representativeness of those individuals, within-patient correlation and the chosen thresholds still matter. Quality results should not be presented as equally secure or generalisable as the full-cohort mortality comparison.

The quality categories are also constructed surrogates. A breath becomes “critical” if any component crosses a threshold, while “optimal” requires a combination of limits. Those limits are not a validated measure of lung or diaphragm injury across all diagnoses. Maximum airway pressure is not interchangeable with plateau or transpulmonary pressure, and an end-tidal carbon dioxide target may have different significance in different clinical settings. Automation can improve conformity to the category definitions without necessarily improving patient outcomes.

Multiplicity is handled more explicitly here than in many trials. The supplement shows that nominal differences in rescue therapy, prone positioning and severe hypoxaemia do not survive correction. It would therefore be inappropriate to construct an adoption case from several individually favourable unadjusted P values. Prone positioning is also a clinician-mediated intervention, not intrinsically an adverse outcome. Fewer prone episodes might reflect better gas exchange, different thresholds or other case-specific decisions; it is not itself proof of less lung injury.

This was a superiority trial. Failure to show superiority does not establish non-inferiority, and similar point estimates do not demonstrate that automation can safely replace staff input. No acceptable clinical margin was prespecified for that purpose. Workload, staffing needs and costs were not measured. Claims that the system achieves the same outcomes with fewer resources are therefore plausible operational hypotheses, not ACTiVE findings. A future non-inferiority evaluation would also need a population and endpoint sensitive enough to detect clinically important loss of quality.

The intervention was one particular platform with a stable algorithm and clinical oversight. It was not a test of all automation, all adaptive ventilation or an independently learning artificial intelligence system. Sensor dependence was sufficiently important to exclude unreliable oximetry, and clinicians retained responsibility for targets and rescue. Extrapolation to other devices, software versions or unsupported autonomous operation would go beyond the evidence.

ACTiVE consequently supports a measured conclusion: the tested system can deliver a distinct ventilation strategy, but its physiological advantages did not translate into more days alive without invasive ventilation in this setting. Selection after randomisation and the limitations of the quality substudy should temper confidence, while the absence of measured resource outcomes prevents a separate claim of demonstrated efficiency.

Body of Evidence

Lellouche and colleagues randomised sixty haemodynamically stable postoperative cardiac patients to automated or protocolised ventilation. Time in a predefined optimal zone was substantially greater with automation, and only five manual interventions were recorded compared with 148 under conventional management. A later 220-patient postoperative trial by de Bie and colleagues similarly improved time within optimal ventilatory limits and accelerated return of spontaneous breathing. These trials support technical and physiological feasibility, but their selected, relatively predictable recovery pathways differ from ACTiVE's medical ICU population.^{4,5}

Arnal and colleagues addressed the operational question in sixty ICU patients expected to need at least 48 hours of ventilation. Median manual setting changes were five versus ten per patient-day with INTELLiVENT-ASV and conventional modes, respectively, while sedation doses and blood-gas sampling did not differ clearly. This is direct evidence of fewer ventilator adjustments in that small study, not proof of reduced total staffing or costs in ACTiVE. Work removed from one task may be replaced by oversight, troubleshooting or other care.⁶

INTELLiPOWER used a randomised crossover design in 96 patients, comparing three-hour periods of automated and conventional ventilation. Overall mechanical power was similar, 15.8 versus 16.1 J/min, with a mean difference of -0.44 J/min (95% CI -1.17 to 0.29). A passive-patient subgroup appeared more favourable, but a short physiological comparison cannot establish a clinical benefit or a validated treatment-selection rule. Even the proposed intermediate mechanism is therefore not uniformly improved across all patients.⁷

A 2025 Cochrane review of automated weaning systems found a probable reduction in ventilation duration across many mostly small trials, but substantial heterogeneity and little clear mortality effect. Devices, populations and comparator protocols varied. Its pooled benefit supplies relevant prior evidence, while ACTiVE tests a more specific early whole-course strategy against an intensive conventional protocol. The two findings should be reconciled through those differences rather than allowing either the meta-analysis or the newest trial to erase the other.⁸

The distinction between improved ventilator settings and improved recovery is also evident in proportional-support research. PROMIZING randomised 573 analysed patients who could tolerate partial support but were not ready for liberation. PAV+ targeted normal estimated respiratory effort, while pressure support was adjusted using respiratory rate and tidal volume. Both arms followed the same structured breathing-trial pathway. Time to sustained liberation was 7.3 versus 6.8 days, with a hazard ratio of 0.96 (95% CI 0.80–1.15). This does not test the ACTiVE algorithm, but it similarly shows that a plausible physiological approach may add little to a strong protocolised comparator.⁹

A 306-patient randomised NAVA trial produced a different result: median ventilator-free days were 22 versus eighteen with conventional ventilation, without a clear mortality difference. NAVA uses diaphragm electrical activity to guide support, and its conventional comparator permitted several modes at clinician discretion. These distinctions prevent treating all adaptive or proportional approaches as one

intervention. A pooled label can obscure clinically important differences in patient selection, timing, technology and background weaning care.¹⁰

Oxygen control is another component for which physiological reduction is not synonymous with benefit. UK-ROX randomised 16,500 patients and reduced supplemental oxygen exposure by approximately 29% under a target SpO₂ of 90%. Ninety-day mortality was 35.4% versus 34.9%, with an adjusted difference of 0.7 percentage points (95% CI -0.7 to 2.0). This separate trial does not determine how an automated system should set oxygen, but cautions against assuming that more conservative oxygenation automatically explains a clinical advantage.¹¹

The appropriate synthesis is therefore specific. ACTiVE adds an important clinical comparison of early comprehensive closed-loop ventilation against expert protocolised care. It does not confirm a patient-centred benefit, nor does it eliminate the possibility of operational value or benefit in a more targeted population. Those remaining questions need direct measurement of outcomes, workload and failures, with retention of the full randomised cohort whenever ethically and legally possible. Improved conformity to ventilator targets alone is insufficient to answer them.

Should early closed-loop ventilation replace protocolised conventional ventilation to improve recovery?

ACTiVE does not support that claim. It found no increase in ventilator-free days and did not establish non-inferiority, reduced workload or cost savings.

References

1. Sinnige JS, Buiteman-Kruizinga LA, Horn J, Paulus F, Schultz MJ, Serpa Neto A. Effect of Automated Closed-Loop Ventilation vs Protocolized Conventional Ventilation on Ventilator-Free Days in Critically Ill Adults: A Randomized Clinical Trial. *JAMA*. 2026;335(10):874-884. Published online December 8, 2025.
2. Sklar MC, Ferguson ND. Automated Modes to Improve Mechanical Ventilation Outcomes—The Ghost in the Machine. *JAMA*. 2025;epublished December 8.
3. Sinnige JS, Buiteman-Kruizinga LA, Horn J, Paulus F, Schultz MJ, Serpa Neto A. Supplement 3 to Effect of Automated Closed-Loop Ventilation vs Protocolized Conventional Ventilation on Ventilator-Free Days in Critically Ill Adults. *JAMA*. 2025.

4. Lellouche F, Bouchard PA, Simard S, L'Her E, Wysocki M. Evaluation of fully automated ventilation: a randomized controlled study in post-cardiac surgery patients. *Intensive Care Med.* 2013;39(3):463-471.
5. De Bie AJR, Neto AS, van Meenen DM, Bouwman AR, Roos AN, Lameijer JR, et al. Fully automated postoperative ventilation in cardiac surgery patients: a randomised clinical trial. *Br J Anaesth.* 2020;125(5):739-749.
6. Arnal JM, Garnero A, Novotni D, Corno G, Donati SY, Demory D, et al. Closed loop ventilation mode in Intensive Care Unit: a randomized controlled clinical trial comparing the numbers of manual ventilator setting changes. *Minerva Anesthesiol.* 2018;84(1):58-67.
7. Buiteman-Kruizinga LA, Serpa Neto A, Botta M, List SS, de Boer BH, van Velzen P, et al. Effect of automated versus conventional ventilation on mechanical power of ventilation-A randomized crossover clinical trial. *PLoS One.* 2024;19(7):e0307155.
8. Rose L, Schultz MJ, Cardwell CR, Paulus F, Couper K, Jouvett P, et al. Automated versus non-automated weaning for reducing the duration of mechanical ventilation for critically ill adults and children. *Cochrane Database Syst Rev.* 2025;7:CD009235.
9. Bosma KJ, Burns KEA, Martin CM, Skrobik Y, Mancebo Cortés J, Mulligan S, et al. Proportional-Assist Ventilation for Minimizing the Duration of Mechanical Ventilation. *N Engl J Med.* 2025;393(11):1088-1103.
10. Kacmarek RM, Villar J, Parrilla D, Alba F, Solano R, Liu S, et al. Neurally adjusted ventilatory assist in acute respiratory failure: a randomized controlled trial. *Intensive Care Med.* 2020;46(12):2327-2337.
11. Martin DS, Gould DW, Shahid T, Doidge JC, Cowden A, Sadique Z, et al. Conservative Oxygen Therapy in Mechanically Ventilated Critically Ill Adult Patients: The UK-ROX Randomized Clinical Trial. *JAMA.* 2025;334(5):398-408.

RESCUE-3

Dres M, Ewert R, Conrad SA, Ataya A, Shrager J, Mortaza S, et al. Temporary Transvenous Diaphragm Neurostimulation for Weaning from Mechanical Ventilation (RESCUE-3): A Randomized Clinical Trial. Am J Respir Crit Care Med. 2026;212(1):86–94. Published online June 11, 2025.

Introduction

Prolonged dependence on mechanical ventilation is often sustained by several interacting problems: unresolved lung disease, cardiovascular limitation, weakness, impaired consciousness and an inability to manage secretions. Diaphragm dysfunction is an attractive treatment target because it directly impairs the respiratory pump and may worsen during ventilation itself. Supporting gas exchange can reduce the activity needed to preserve respiratory muscle function, while critical illness adds inflammatory, metabolic and neurological injury.

Conventional inspiratory muscle training asks patients to breathe against an imposed load. It depends to varying degrees on cooperation and the ability to generate useful effort. Transvenous diaphragm neurostimulation takes a different approach. Electrodes incorporated into a central venous catheter stimulate the phrenic nerves, producing diaphragm contractions even when voluntary training is difficult. The intended treatment is rehabilitation of the respiratory muscles, rather than continuous replacement of ventilatory support.¹

Earlier studies established technical feasibility and suggested improvements in maximal inspiratory pressure. Whether those changes produce more successful liberation is a separate question. A stronger inspiratory effort may not overcome cardiac failure, airway problems or persistent neurological impairment. Moreover, inserting and maintaining an additional central catheter introduces risks that must be considered alongside any reduction in ventilation.

RESCUE-3 tested this strategy in patients with established difficulty weaning. Its interpretation is complicated by early administrative stopping, exclusion of unsuccessful device recipients from the primary population and a Bayesian analysis incorporating information from RESCUE-2. The study provides an encouraging signal, but the strength and independence of that signal need to remain visible.

Synopsis

RESCUE-3 was an international, open-label randomised trial conducted at 48 centres in the United States and Europe. Recruitment ran from June 2019 to January 2023. Lungpacer Medical, the device manufacturer, supported the study and participated in its design, analysis and reporting. An independent data safety monitoring board reviewed safety and interim results; an unblinded clinical events committee adjudicated serious events. The report appeared online on 11 June 2025.^{1,2}

Adults were eligible after at least four days of invasive ventilation if they met readiness-to-wean criteria and had failed at least two weaning attempts on different days. At least one failure had to occur under standardised low-support conditions. Important exclusions included an absolute maximal inspiratory pressure above 50 cm H₂O, ventilation for more than ninety days, extracorporeal support, weaning failure attributed to volume overload, contraindications to the required venous access and implanted electrical devices potentially affected by stimulation.

Central electronic randomisation allocated patients equally to stimulation plus standard care or standard care alone, using variable blocks and stratification by centre. Both groups received protocolised daily weaning assessments when eligible, sedation minimisation and mobilisation according to local practice. Separate inspiratory muscle training protocols were prohibited. Non-invasive ventilation or high-flow oxygen could be used after liberation to prevent respiratory failure.

The intervention required a multielectrode catheter inserted through the left subclavian or internal jugular vein and positioned near the cavoatrial junction. Two device generations were used; the later system added jugular access and workflow modifications while retaining equivalent stimulation characteristics. Treatment comprised two daily sessions, each containing six sets of ten stimulations, separated by at least three hours. Intensity was increased according to tolerance and clinical judgement. Treatment continued for up to thirty days, successful weaning, death or withdrawal.

The primary outcome was successful weaning by day 30, defined as passing a standardised trial and remaining off mechanical ventilation for at least 48 hours. Secondary outcomes included ventilation duration, change in maximal inspiratory pressure, rapid shallow breathing index and mortality. Exploratory outcomes included sustained liberation for seven days, ventilator-free days, reconnection to ventilation, tracheostomy and length of stay.

The planned Bayesian adaptive design allowed recruitment up to 400 patients, with prespecified interim thresholds for declaring superiority. A subset of 85 RESCUE-2 participants meeting the new eligibility criteria supplied historical information. Propensity-based weighting, using clinical characteristics relevant to outcome, reduced their contribution to the equivalent of 64 patients. The primary time-to-weaning model combined that prior information with the new trial data.

Patients who died before day 30 were treated as unsuccessful weaning observations through day 30, rather than being counted as successfully liberated by death. Withdrawals were censored when follow-up ended. The primary modified intention-to-treat population excluded intervention patients in whom catheter placement or effective stimulation could not be achieved, as well as those who withdrew before the procedure. Safety analyses included all randomised patients.

At the interim review with approximately 200 patients, the probability of superiority was 92.7%, below the prespecified 99.25% stopping threshold. The estimated conditional power to reach the superiority criterion at maximum recruitment was 44.6%. Designated sponsor representatives, and subsequently its board, reviewed these results. Facing slow recruitment and the cost of continuing towards 400 participants, the sponsor stopped the trial. Recruitment during the interim analysis had brought the final randomised total to 223.²

There were 109 intervention and 114 control allocations. Five intervention patients had unsuccessful placement or nerve capture, and two withdrew before the procedure, leaving 102 versus 114 in the primary analysis. Patients had received ventilation for a mean of 29 days, 64% already had a tracheostomy, and baseline maximal inspiratory pressure averaged approximately 29 cm H₂O. This was therefore predominantly a prolonged-weaning population, despite eligibility beginning after four days.

Intervention patients received approximately 79% of scheduled stimulations, and 88% received at least half the required dose. Successful weaning occurred in 71/102 patients (69.6%) versus 69/114 (60.5%). With historical borrowing, the adjusted hazard ratio was 1.34 (95% credible interval 1.01–1.78), with a 97.9% posterior probability of superiority. Without borrowing, the corresponding hazard ratio was 1.26 (95% credible interval 0.91–1.75) and the probability was 91.5%.

The estimated reduction in ventilation duration was approximately 2.5 days, with a 95% credible interval from 5.0 days shorter to 0.1 days longer. Maximal inspiratory pressure improved more with stimulation, with an adjusted difference of 5.9 cm H₂O (95% credible interval 1.8–10.0). A seven-day definition of successful liberation also favoured

treatment. Mortality was 10/102 (9.8%) versus 12/114 (10.5%), with an adjusted hazard ratio of 0.74 (95% credible interval 0.37–1.46).

Serious adverse events occurred in 39/109 intervention patients (35.8%) and 27/114 controls (23.7%). Two intervention-group deaths followed events adjudicated as possibly related to the device or insertion procedure, involving haemothorax or possible tension pneumothorax. Stimulation itself caused little reported additional pain. These findings distinguish procedural and device safety from immediate tolerability of an individual stimulation session.^{1,2}

Critique

RESCUE-3 supports further investigation of diaphragm neurostimulation, but does not provide an uncomplicated confirmatory result. The observed weaning proportions and physiological response point in a favourable direction. However, the trial stopped before reaching its planned superiority criterion, the final estimate depends partly on historical data, and serious adverse events were more common. Each limitation bears directly on whether the intervention should become routine care.

The Bayesian probability is interpretable only in relation to the model and prior information. A 97.9% probability that the hazard ratio exceeds one is not a 97.9% probability that every patient benefits, that the treatment is safe, or that the effect is large enough to justify its costs and complications. It also does not establish the probability of a specified clinically important absolute benefit. The raw difference in weaning proportions was approximately nine percentage points, but uncertainty in that difference remains important.

Historical borrowing strengthened the result materially. The new trial alone produced a credible interval crossing no effect and a probability of superiority of 91.5%. Borrowing is not inherently inappropriate: the studies addressed closely related populations and interventions, and the approach was planned. Nevertheless, downweighting measured differences cannot guarantee exchangeability across trials. Recruitment periods, devices, background care and unmeasured prognostic features may differ even after matching on recorded variables.

The earlier trial must therefore not be counted twice in a narrative synthesis. It contributes both to the history supporting RESCUE-3 and to the final Bayesian estimate. Describing two independent positive trials would be incorrect: RESCUE-2 did not establish improved weaning, and part of its information is already embedded in the

RESCUE-3 result. The unborrowed analysis is especially useful because it shows what the new randomised evidence contributes on its own.

Early stopping deserves a precise account. Recruitment was curtailed for financial and practical reasons, but the decision was informed by unblinded interim efficacy results and the estimated prospect of eventual success. It was not a wholly data-independent interruption. The sponsor's decision does not make Bayesian estimation impossible, and a final posterior can still summarise the observed information under the model. It does, however, leave a smaller and less precise clinical evidence base than planned and does not retrospectively satisfy the original decision rule.

The distinction between estimation and declaration is central. A final credible interval narrowly excluding one can coexist with failure to cross a more demanding prespecified stopping boundary. The appropriate account is an encouraging posterior estimate from an incompletely recruited trial. Calling the study definitively positive solely because the 95% interval excludes one discards part of its design; calling the data uninterpretable solely because recruitment ended early discards useful information.

The modified intention-to-treat population also favours a question narrower than routine implementation. A clinician deciding to offer the device must include the possibility of failed access, failed nerve capture and deterioration before treatment. Excluding such patients after allocation weakens randomisation and estimates benefit among those in whom delivery was possible. Five unsuccessful recipients had follow-up discontinued after the short safety window, so their day-30 outcomes were not simply available for a complete allocation-based analysis.

Sensitivity analyses imputed possible outcomes for those patients and retained a generally favourable direction, with probabilities of superiority between 94.5% and 98.8%. This is useful, but the probabilities assigned to different missing-outcome scenarios are assumptions rather than observed follow-up. A per-protocol analysis is even more selected because adequate treatment delivery may relate to prognosis. Higher certainty in that population cannot restore the missing intention-to-treat comparison.

Open treatment allocation creates further uncertainty. Daily standardised weaning criteria and similar reported sedation and ventilator management reduce opportunities for differential care. They do not eliminate differences in encouragement, readiness assessment, mobilisation or decisions to attempt liberation. Maximal inspiratory pressure also depends on measurement conditions and patient performance. A coherent

improvement in strength and weaning supports the proposed mechanism, but neither outcome is completely insulated from knowledge of treatment.

The mechanism itself remains incompletely demonstrated. Maximal inspiratory pressure reflects the combined inspiratory muscles rather than diaphragm contractility alone. Both groups improved, and the between-group difference was approximately 6 cm H₂O. The accompanying editorial questions whether that increment fully explains the apparent clinical effect. It may contribute, but mediation was not established, and changes in respiratory coordination, lung recruitment or other care cannot be separated by the available data.³

Safety cannot be dismissed as ascertainment bias. Unblinded staff may report more events in device recipients, and the events committee was also unblinded. Yet the intervention necessarily adds an invasive catheter, whereas controls need not have equivalent access. Catheter infection, vascular injury and pneumothorax are therefore part of the strategy's potential harm, not nuisance differences to remove from the comparison. The two possibly related fatal events make a broad description of the device as proven safe inappropriate.

Nor does the small numerical mortality difference settle the balance. Only 22 deaths occurred in the primary population, and the mortality credible interval includes substantial benefit and harm. Serious adverse events were analysed in the larger randomised safety population, so their denominators differ from those for efficacy and mortality. These populations should be stated explicitly, particularly when safety and benefit are presented together.

The endpoint is meaningful but limited. Forty-eight hours off ventilation may be followed by reconnection; the favourable seven-day sensitivity analysis helps, without establishing sustained independence after discharge. Thirty-day follow-up does not show that earlier liberation improves long-term function, quality of life or survival. No clear reduction in ICU or hospital stay was established. Resource savings remain plausible, rather than directly demonstrated consequences of the trial.

Applicability is also restricted. Most participants had already spent weeks ventilated, and many had tracheostomies. They met readiness criteria and had demonstrable weaning difficulty with limited inspiratory pressure. The study does not justify preventive implantation in newly intubated patients, stimulation during severe unstable respiratory failure or treatment of every failed spontaneous breathing trial. Excluding volume-overload-related failure appropriately recognises that respiratory weakness is only one possible barrier to liberation.

Finally, manufacturer involvement and the stated restriction on public access to proprietary participant data make transparent reporting particularly important. Commercial support does not invalidate the randomised comparison. It does increase the value of independent replication, complete allocation-based follow-up and explicit reporting of procedure-related complications. A future study should resolve the clinical benefit–harm balance with adequate recruitment and patient-centred outcomes beyond the first month.

Body of Evidence

Human diaphragm biopsy research supplied a strong biological rationale for avoiding prolonged inactivity. Levine and colleagues compared fourteen brain-dead organ donors ventilated with inactive diaphragms for 18–69 hours with eight surgical controls exposed for only a few hours. Marked fibre atrophy and biochemical evidence consistent with increased proteolysis were observed. This establishes biological plausibility, but an observational comparison involving brain death cannot by itself show that a particular stimulation strategy improves outcomes in general ICU patients.⁴

RESCUE-1 was a small feasibility study. Eleven patients met entry criteria, nine underwent successful catheter placement, and stimulation produced diaphragm contractions in all nine. Seven met weaning criteria. Without a concurrent control group, spontaneous recovery and other treatment cannot be separated from device effects. The study's main contribution was showing that transvenous phrenic stimulation could be delivered in selected patients, rather than establishing efficacy.⁵

RESCUE-2 then randomised 112 patients, allocating 57 to stimulation and 55 to usual care. Catheter insertion or delivery failed in fourteen intervention patients, leaving data for 43 treated recipients. Successful weaning was 82% versus 74%, with an estimated absolute difference of seven percentage points (95% CI –10 to 25), and ventilation duration did not differ significantly. Maximal inspiratory pressure improved more with stimulation, with a between-group difference of 11.8 cm H₂O (95% CI 5–19). Serious adverse event frequency was similar.⁶

That pattern suggested a physiological effect without a secure clinical effect, and also exposed a substantial delivery problem. RESCUE-3 appears to have improved technical feasibility, but retained a primary analysis excluding unsuccessful recipients. The consistency in the direction of weaning results is encouraging, while the dependence created by historical borrowing and the different safety signal prevent treating the sequence as definitive independent replication.

IMweanT provides a useful comparison with conventional loading-based rehabilitation. Ninety difficult-to-wean patients were randomised to high-intensity inspiratory muscle training or a low-intensity sham approach. Weaning success was 64% versus 76%, without a significant difference, and both groups improved maximal inspiratory pressure by approximately 14–15 cm H₂O. The trial illustrates how respiratory strength can improve during recovery without a corresponding advantage in liberation from a more intensive training prescription.⁷

A 2025 systematic review of eighteen inspiratory muscle training trials found very uncertain evidence about physical function, with only three studies reporting that primary outcome. Different training methods should not be assumed to have the same effects as electrical stimulation, but the review reinforces a broader gap: evidence of increased strength is more available than evidence of improved long-term recovery. RESCUE-3 advances the clinical question, while leaving that gap largely unresolved.⁸

The current evidence consequently supports a promising rehabilitation concept with incomplete confirmation of net clinical benefit. The next decisive study should preserve all randomised patients, minimise bias in liberation assessment, retain transparent safety adjudication and measure sustained recovery. RESCUE-3 is an important step towards that evidence, rather than a sufficient basis for routine device use across prolonged ventilation.

Should transvenous diaphragm neurostimulation become routine treatment for difficult weaning?

RESCUE-3 provides a promising but uncertain weaning signal, strengthened by borrowed data and accompanied by more serious adverse events. It does not yet establish a sufficiently secure benefit–harm balance for routine use.

References

1. Dres M, Ewert R, Conrad SA, Ataya A, Shrager J, Mortaza S, et al. Temporary Transvenous Diaphragm Neurostimulation for Weaning from Mechanical Ventilation (RESCUE-3): A Randomized Clinical Trial. *Am J Respir Crit Care Med*. 2026;212(1):86-94.
2. Dres M, Ewert R, Conrad SA, Ataya A, Shrager J, Mortaza S, et al. RESCUE-3 online data supplement. *Am J Respir Crit Care Med*. Published online 2025. doi:10.1164/rccm.202505-1056OC.

3. Mauri T, Heunks L. Diaphragm Neurostimulation: An Innovative Approach to Treat Diaphragm Weakness in Difficult to Wean Patients? *Am J Respir Crit Care Med*. 2026;212(1):11–12. Published online July 28, 2025.
4. Levine S, Nguyen T, Taylor N, Friscia ME, Budak MT, Rothenberg P, et al. Rapid disuse atrophy of diaphragm fibers in mechanically ventilated humans. *N Engl J Med*. 2008;358(13):1327-1335.
5. Ataya A, Silverman EP, Bagchi A, Sarwal A, Criner GJ, McDonagh DL. Temporary Transvenous Diaphragmatic Neurostimulation in Prolonged Mechanically Ventilated Patients: A Feasibility Trial (RESCUE 1). *Crit Care Explor*. 2020;2(4):e0106.
6. Dres M, de Abreu MG, Merdji H, Müller-Redetzky H, Dellweg D, Randerath WJ, et al. Randomized Clinical Study of Temporary Transvenous Phrenic Nerve Stimulation in Difficult-to-Wean Patients. *Am J Respir Crit Care Med*. 2022;205(10):1169-1178.
7. Van Hollebeke M, Poddighe D, Hoffman M, Clerckx B, Muller J, Louvaris Z, et al. Similar Weaning Success Rate with High-Intensity and Sham Inspiratory Muscle Training: A Randomized Controlled Trial (IMweanT). *Am J Respir Crit Care Med*. 2025;211(3):381-390.
8. Farley C, Oliveira A, Brooks D, Newman ANL. The Effects of Inspiratory Muscle Training in Critically ill Adults: A Systematic Review and Meta-Analysis. *J Intensive Care Med*. 2026;41(4):263-273. Published online February 9, 2025.

DESIGNATION

**Writing and Steering Committees for the DESIGNATION–Investigators.
Intraoperative Driving Pressure-Guided High PEEP vs Standard Low PEEP for
Postoperative Pulmonary Complications. JAMA. 2026;335(8):693-702.**

Introduction

The attraction of individualised ventilation is straightforward: patients differ in lung size, chest-wall mechanics and the amount of recruitable atelectasis, so a single positive end-expiratory pressure may be unsuitable for everyone. The difficult question is how to choose a setting that improves recovery, rather than simply improving a number on the ventilator. During abdominal surgery, recruitment and higher positive end-expiratory pressure may reopen dependent lung while simultaneously reducing venous return and increasing the circulatory burden of anaesthesia.

Driving pressure, calculated as plateau pressure minus positive end-expiratory pressure, offers an apparently practical guide. At a fixed tidal volume, a lower driving pressure indicates greater respiratory-system compliance. An individual-patient meta-analysis of surgical ventilation trials associated higher driving pressure with postoperative pulmonary complications. However, patients were not randomised to their measured driving pressures, and an association between a physiological variable and outcome does not establish that manipulating the variable improves outcome.¹

Earlier trials established benefits from protective ventilation bundles compared with relatively injurious ventilation, but trials specifically increasing positive end-expiratory pressure produced less encouraging results. A small trial suggested that choosing positive end-expiratory pressure at the lowest driving pressure might overcome the limitations of applying the same higher pressure to everyone.^{2,3,4,5}

DESIGNATION tested this proposition in a large international population undergoing open abdominal surgery. Its importance lies in separating successful manipulation of lung mechanics from clinical effectiveness. The trial asked whether repeated recruitment manoeuvres and driving-pressure-guided higher positive end-expiratory pressure prevented pulmonary complications when both groups already received relatively low tidal volumes. Although its journal issue appeared in 2026, the report was first published online on 3 December 2025.⁶

Synopsis

DESIGNATION was an investigator-initiated, assessor-blinded, parallel-group superiority trial in twenty-nine hospitals across five European countries. Recruitment ran from April 2019 to December 2024. The Dutch health research organisation ZonMw funded the trial without a reported role in its conduct or interpretation. Participants gave written informed consent, and an independent monitoring board reviewed safety. Treating clinicians necessarily knew the ventilator strategy; postoperative outcome assessors were masked to allocation.⁶

Eligible adults required open abdominal surgery and an ARISCAT score of at least 26, indicating intermediate or high predicted pulmonary risk. Important exclusions included laparoscopic or robotic surgery, combined intrathoracic procedures, prone or lateral positioning, body mass index above 40 kg/m², substantial recent mechanical ventilation and expected postoperative invasive ventilation. Patients with shock, severe cardiac disease or severe chronic respiratory disease were also excluded. The comparison therefore concerned perioperative prevention in selected surgical patients, rather than ventilation of patients with established critical respiratory failure.

Randomisation used concealed allocation in equal proportions, with permuted blocks and stratification by centre and body mass index. Separate intraoperative and postoperative data collection helped maintain assessor masking. Both groups received volume-controlled ventilation with low tidal volumes based on predicted body weight. Actual median tidal volumes were approximately 8 mL/kg in each group, an important detail when describing the background ventilation.

The intervention combined recruitment manoeuvres with individualised higher positive end-expiratory pressure. Once haemodynamics were stable, pressure was increased from 5 to 20 cm H₂O in 5-cm steps during recruitment. A decremental trial then reduced positive end-expiratory pressure from 20 to 6 cm H₂O in 2-cm steps. Clinicians selected the highest positive end-expiratory pressure associated with the lowest driving pressure. If no minimum was identified, the default setting was 12 cm H₂O. Recruitment was repeated around titration, after circuit disconnection and within the final hour before extubation.^{6,7}

Control patients received positive end-expiratory pressure of 5 cm H₂O without routine recruitment or driving-pressure titration. Rescue adjustments were permitted, and safety rules allowed pressure reductions for haemodynamic instability or other serious concerns. Anaesthesia, fluids, analgesia, reversal of neuromuscular blockade and postoperative care largely followed local practice. Enhanced recovery was encouraged rather than delivered as a uniform trial intervention.

The primary endpoint was at least one of ten pulmonary complications during the first five postoperative days. These included severe respiratory failure, bronchospasm, suspected pulmonary infection, new pulmonary infiltrates, aspiration, atelectasis, acute respiratory distress syndrome, pleural effusion, cardiopulmonary oedema and pneumothorax. Secondary endpoints addressed individual pulmonary events, other complications, intensive care admission, hospital stay and hospital mortality. Intraoperative safety outcomes included hypotension and desaturation. Hypotension meant a mean arterial pressure more than 20% below baseline for over three minutes; desaturation meant oxygen saturation below 92% for over one minute.

The planned sample of 1,468 provided 80% power to detect a relative reduction of 20%, from an assumed complication risk of 34% to 27.2%, allowing for withdrawals. The primary analysis used a mixed-effects model with centre as a random effect and reported an adjusted absolute risk difference. It was a modified intention-to-treat analysis, excluding patients who withdrew or became ineligible before receiving the allocated strategy. Holm adjustment addressed eighteen secondary comparisons. The primary comparison itself did not require that secondary-outcome adjustment.^{6,7}

Of 7,446 patients screened, 1,468 were randomised, 734 per group. Thirty-three were excluded before treatment, leaving 718 intervention and 717 control participants. Changes to laparoscopic surgery accounted for most exclusions. Participants were typically around 66 years old, 52% were women, and median body mass index was approximately 25 kg/m². Most procedures were elective, approximately 71% of participants had active cancer, and only about one-fifth were in the highest ARISCAT risk category.

The intervention produced clear physiological separation. Almost every intervention patient received recruitment manoeuvres, with a median of three, whereas routine recruitment was essentially absent in controls. After one hour, median positive end-expiratory pressure was 10 versus 5 cm H₂O and driving pressure was 8 versus 10 cm H₂O. Plateau pressure was higher, at 19 versus 15 cm H₂O. Tidal volume and inspired oxygen concentration were similar. No driving-pressure minimum was identified in approximately 10% of intervention patients, who therefore received the default pressure. Safety-related pressure adjustments were more frequent with the intervention.^{6,7}

The primary endpoint occurred in 142/718 intervention patients, 19.8%, and 125/717 controls, 17.4%. The adjusted absolute difference was 2.5 percentage points, with a 95% confidence interval from -1.5 to 6.4; P=0.23. Atelectasis and pleural effusion were the commonest components. Severe respiratory failure occurred in thirteen versus ten

patients and acute respiratory distress syndrome in four versus two. Individual components could overlap within the same patient.⁶

Hypotension occurred in 382/708 patients with available data in the intervention group, 54.0%, versus 317/704 controls, 45.0%. The absolute difference was 9.0 percentage points, 95% confidence interval 4.7–13.4, and remained significant after multiplicity adjustment. The recorded vasoactive-agent complication endpoint also favoured control: 224/700, 32.0%, versus 130/690, 18.8%. This endpoint should not be interpreted as the proportion receiving any vasoactive drug at any time, because the separate medication table recorded much more frequent norepinephrine exposure in both groups.^{6,7}

Desaturation occurred in six versus twenty patients, an absolute difference of –2.0 percentage points. Its unadjusted P value was 0.005, but the Holm-adjusted value was 0.08. Median hospital stay was six days in both groups. Hospital mortality was 25/698 versus 23/703, approximately 3.6% versus 3.3%, with no demonstrated difference; mortality follow-up was unavailable for thirty-four patients. Per-protocol and additional adjusted analyses did not establish pulmonary benefit.^{6,7}

Critique

The trial provides a credible answer to a clinically relevant question because its control strategy resembles contemporary protective ventilation. Both groups received similar tidal volumes and the control group had modest positive end-expiratory pressure. The contrast was consequently the added value of a specific recruitment-and-titration strategy. It was not a comparison between comprehensive protective ventilation and an intentionally less protective regimen. Concealed randomisation, broad hospital participation, masked postoperative assessment and complete primary-endpoint ascertainment in the analysed cohort strengthen the result.

Intervention fidelity also supports interpretation. Recruitment manoeuvres, positive end-expiratory pressure and driving pressure differed substantially between groups. Failure to improve the primary endpoint cannot readily be explained by failure to alter the targeted physiology. Instead, DESIGNATION demonstrates the gap between a plausible intermediate target and an effective treatment strategy. Lower driving pressure at a fixed tidal volume can coexist with higher plateau pressure and greater haemodynamic disturbance. It does not follow that every way of lowering driving pressure has the same biological consequences.

The strategy nevertheless represents only one form of individualisation. Pressure titration followed recruitment, could not select a pressure below 6 cm H₂O, favoured the

highest setting when several settings produced the same minimum, and used 12 cm H₂O when no minimum emerged. These choices encouraged higher pressures. They did not test a flexible bedside approach that also altered tidal volume, accepted lower pressure or incorporated cardiac output and recruitability into the decision. The negative result applies directly to this bundled algorithm; extrapolating it to every personalised ventilation strategy would be excessive.^{6,7}

The unexpectedly low control event rate reduced statistical power for the originally anticipated relative treatment effect. Only 17.4% of controls had a primary event, compared with the assumed 34%. However, this does not make the trial uninformative. The primary confidence interval was compatible with at most approximately 1.5 percentage points of benefit and up to 6.4 points of harm. It excluded the planned 6.8-point absolute benefit. A small benefit remains possible, but a large preventive effect in this population is difficult to reconcile with the data.

The composite endpoint requires attention to severity. Atelectasis and effusions were much more common than acute respiratory distress syndrome or severe respiratory failure. Equal weighting of these events means that the primary endpoint does not directly measure avoidance of life-threatening respiratory deterioration. Conversely, the absence of a primary benefit was not concealed by a persuasive improvement in a severe component. Event counts for individual serious complications were simply too small for reliable conclusions. Neither mortality nor rare pulmonary outcomes establish equivalence between strategies.

Blinding was strongest where it mattered for postoperative ascertainment, but remained impossible for intraoperative care. Clinicians could change fluids, vasoactive treatment and ventilator settings after observing the assigned intervention. Those responses are part of the practical consequences of implementing the strategy, although clinician discretion complicates interpretation of treatment-dependent endpoints. A measured hypotension episode is less discretionary than the decision to administer a drug. The consistency of lower blood pressure tolerance with higher pressure requirements makes the hypotension signal clinically persuasive without establishing the precise mechanism in each patient.

The vasoactive-agent results also contain a reporting ambiguity. The main outcome table gives rates of 32.0% and 18.8% for its recorded endpoint, while the supplementary medication table records norepinephrine use in approximately 85% of both groups. These cannot both describe unrestricted exposure to any vasoactive drug. The distinction may depend on the endpoint's detailed operational definition, but the reviewed results do not resolve it. The chapter therefore retains the numerical result

with that qualification and gives greater interpretative weight to the clearly specified hypotension endpoint.^{6,7}

Multiplicity changes the oxygenation story. Fewer intervention patients desaturated, but this comparison did not meet the trial's adjusted significance criterion. Describing it as definitively improved while applying stricter standards to unfavourable outcomes would be inconsistent. By contrast, hypotension and the recorded vasoactive endpoint remained significant after adjustment. The evidence therefore supports a haemodynamic cost more securely than it supports a clinically meaningful oxygenation benefit.

The modified intention-to-treat population introduces a modest additional qualification. Thirty-three participants were excluded after randomisation, mainly because the planned procedure changed before treatment. Numbers were small and balanced, reducing concern, but the analysis was not literally an analysis of every randomised participant. The supplementary per-protocol table also uses different denominators from the main flow diagram. These analyses broadly support the primary conclusion, but their discrepant populations should not be silently treated as interchangeable. Complete primary follow-up likewise does not imply complete mortality follow-up.

Several additional analyses merit restraint. An exploratory model combining correlated pulmonary components and an analysis of the average number of complications produced results suggesting harm. They do not replace the prespecified primary endpoint. A subgroup with planned postoperative intensive or high-dependency care also appeared to fare worse with the intervention, with an interaction P value of 0.007. This is a useful hypothesis about more vulnerable patients, not a validated rule for selecting treatment. Similarly, the reported confidence-distribution probability of benefit is not a Bayesian posterior probability and does not constitute an independent replication.^{6,7}

Generalisability is strongest for predominantly elective open abdominal surgery in patients without severe obesity, advanced cardiopulmonary disease or anticipated postoperative ventilation. The trial cannot settle pressure selection during laparoscopy, one-lung ventilation, established acute respiratory distress syndrome or shock. Within its studied population, however, repeated recruitment and manual titration added work and circulatory disturbance without a demonstrated postoperative benefit. Routine implementation therefore requires a stronger justification than improvement in driving pressure alone.

Body of Evidence

The original IMPROVE trial established why the background comparator matters. Four hundred patients at increased pulmonary risk undergoing major abdominal surgery were randomised to protective or non-protective ventilation. Its seven-day composite of major pulmonary and extrapulmonary complications occurred in 10.5% versus 27.5%, relative risk 0.40, 95% confidence interval 0.24–0.68. The intervention combined several changes, including lower tidal volumes and positive end-expiratory pressure with recruitment. It supported a protective bundle, but could not identify higher positive end-expiratory pressure as the component responsible for benefit. DESIGNATION builds on that foundation by providing similar tidal volumes in both groups.²

PROVHILO then examined the pressure question more directly. Nine hundred patients undergoing open abdominal surgery were randomised to positive end-expiratory pressure of 12 cm H₂O with recruitment or low pressure without routine recruitment; tidal volumes were approximately 8 mL/kg in both groups. Pulmonary complications occurred in 174/445 and 172/449 evaluable patients, approximately 40% and 39%, with no benefit from higher pressure. Hypotension and vasoactive treatment were more frequent. DESIGNATION reproduces the central clinical pattern despite replacing a fixed higher setting with a physiological titration procedure.³

PROBESE tested a related strategy in patients with body mass index of at least 35 kg/m², who might be expected to have more recruitable dependent collapse. Among 1,976 analysed participants, higher pressure with recruitment produced pulmonary complications in 21.3%, compared with 23.6% with lower pressure. The absolute difference was –2.3 percentage points, 95% confidence interval –5.9 to 1.4. Thus, even in obesity, improved intraoperative respiratory conditions did not establish a reduction in postoperative pulmonary complications. Its population differs from DESIGNATION, which excluded body mass index above 40 kg/m², but the findings weaken the assumption that a physiological rationale alone identifies a benefiting population.⁴

The driving-pressure individual-patient meta-analysis provided the more specific rationale for DESIGNATION. Across seventeen randomised trials involving 2,250 participants, higher driving pressure was associated with complications, with an odds ratio of 1.16 per unit increase. Its mediation analyses suggested that changes in driving pressure could help explain differences between ventilation strategies. Nevertheless, the measured pressure was a postrandomisation physiological variable, affected by patient characteristics and treatment. The analysis generated a hypothesis for a randomised intervention; it did not validate driving-pressure reduction as a surrogate endpoint.¹

A single-centre trial by Zhang and colleagues subsequently randomised 148 patients undergoing open upper abdominal surgery. In its analysed population of 134, clinically significant pulmonary complications occurred in 26/67 patients receiving driving-pressure-guided pressure selection and 42/67 receiving fixed positive end-expiratory pressure of 6 cm H₂O: 38.8% versus 62.7%. The encouraging result was accompanied by better compliance and less atelectasis. However, the small sample, selected setting and much higher control event rate limited its ability to establish routine practice. DESIGNATION provides a substantially larger, multicentre test and fails to confirm a comparable benefit.⁵

iPROVE considered a broader perioperative pathway, combining individualised intraoperative ventilation with different postoperative continuous positive airway pressure strategies. Its four-arm comparison included 967 analysed patients. The primary seven-day composite of pulmonary and systemic complications occurred in approximately 46–48% of the intervention groups and 51% of the standard-ventilation, oxygen-therapy group; none of the primary comparisons showed a statistically significant reduction. It should therefore not be cited as a straightforward positive trial of individualised pressure. It also addresses a different treatment package from DESIGNATION's predominantly intraoperative intervention.⁸

Taken together, these trials support low-tidal-volume protective ventilation while challenging routine escalation of positive end-expiratory pressure with recruitment merely because compliance or oxygenation improves. DESIGNATION adds the crucial observation that selecting higher pressure by the minimum driving pressure does not, by itself, overcome the limitations seen with fixed higher-pressure strategies. The remaining question is whether a better-defined physiological phenotype or a different algorithm can identify benefit without disproportionate haemodynamic cost.

Should driving-pressure-guided higher PEEP with recruitment manoeuvres routinely replace PEEP of 5 cm H₂O during open abdominal surgery?

No. In the population studied, this strategy did not reduce postoperative pulmonary complications and increased intraoperative hypotension despite improving lung mechanics.

References

1. Neto AS, Hemmes SN, Barbas CS, Beiderlinden M, Fernandez-Bustamante A, Futier E, et al. Association between driving pressure and development of postoperative pulmonary complications in patients undergoing mechanical ventilation for general anaesthesia: a meta-analysis of individual patient data. *Lancet Respir Med*. 2016;4(4):272-280.
2. Futier E, Constantin JM, Paugam-Burtz C, Pascal J, Eurin M, Neuschwander A, et al. A trial of intraoperative low-tidal-volume ventilation in abdominal surgery. *N Engl J Med*. 2013;369(5):428-437.
3. Hemmes SN, Gama de Abreu M, Pelosi P, Schultz MJ. High versus low positive end-expiratory pressure during general anaesthesia for open abdominal surgery (PROVHILO trial): a multicentre randomised controlled trial. *Lancet*. 2014;384(9942):495-503.
4. Writing Committee for the PROBESE Collaborative Group of the PROtective VEntilation Network (PROVENet) for the Clinical Trial Network of the European Society of Anaesthesiology. Effect of Intraoperative High Positive End-Expiratory Pressure (PEEP) With Recruitment Maneuvers vs Low PEEP on Postoperative Pulmonary Complications in Obese Patients: A Randomized Clinical Trial. *JAMA*. 2019;321(23):2292-2305.
5. Zhang C, Xu F, Li W, Tong X, Xia R, Wang W, et al. Driving Pressure-Guided Individualized Positive End-Expiratory Pressure in Abdominal Surgery: A Randomized Controlled Trial. *Anesth Analg*. 2021;133(5):1197-1205.
6. Writing and Steering Committees for the DESIGNATION–Investigators. Intraoperative Driving Pressure-Guided High PEEP vs Standard Low PEEP for Postoperative Pulmonary Complications. *JAMA*. 2026;335(8):693-702.
7. Writing and Steering Committees for the DESIGNATION–Investigators. Supplementary methods and results to Intraoperative Driving Pressure–Guided High PEEP vs Standard Low PEEP for Postoperative Pulmonary Complications. *JAMA*. First published online 3 December 2025.
8. Ferrando C, Soro M, Unzueta C, Suarez-Sipmann F, Canet J, Librero J, et al. Individualised perioperative open-lung approach versus standard protective ventilation in abdominal surgery (iPROVE): a randomised controlled trial. *Lancet Respir Med*. 2018;6(3):193-203.

NAVIGATE

Monti G, Cabrini L, Kotani Y, Brusasco C, Kadralinova A, Giardina G, et al. Early noninvasive ventilation in general wards for acute respiratory failure: an international, multicentre, open-label, randomised trial. Br J Anaesth. 2025;134(2):382-391.

Introduction

Acute respiratory deterioration often begins on a general ward, before a patient needs intensive care. Treating the cause and providing oxygen may be sufficient, but some patients progress to marked distress, worsening gas exchange and a need for more advanced support. Intervening before that progression is attractive, provided the treatment itself does not impose disproportionate discomfort, workload or risk.

Non-invasive ventilation has an established role in selected respiratory syndromes. A landmark ward-based trial in exacerbations of chronic obstructive pulmonary disease showed that early ventilatory assistance reduced treatment failure and in-hospital mortality. That evidence cannot automatically be transferred to predominantly hypoxaemic respiratory failure caused by pneumonia, surgery or a mixture of diagnoses. The physiology, treatment responsiveness and consequences of delayed escalation differ.¹

Small studies in haematology wards provided conflicting evidence for early non-invasive support outside intensive care. One suggested substantial reductions in ICU admission and intubation; another found no clear improvement in these outcomes or survival. Their narrow populations and small samples left uncertainty about routine early treatment of the broader deteriorating ward population.^{2,3}

NAVIGATE examined scheduled, intermittent non-invasive support in adults with mild acute respiratory failure who did not initially require intensive care. Most received continuous positive airway pressure (CPAP), which supplies a constant distending pressure without additional inspiratory assistance. The hypothesis was that beginning this treatment before severe deterioration would reduce progression to severe respiratory failure compared with usual care and rescue support when needed. The distinction between preventing a physiological threshold and preventing intubation or death is central to interpreting the result.⁴

Synopsis

NAVIGATE was an international, open-label, parallel-group superiority trial at 11 hospitals in Italy, Greece and Kazakhstan. Recruitment extended from 6 May 2012 to 18 July 2023. An independent epidemiologist generated a computerised, permuted-block allocation sequence stratified by centre. A web system concealed assignment until the required participant information had been entered. Patients or legal representatives provided written consent, and a data and safety monitoring board oversaw the trial. The principal report did not provide a specific funding statement.⁴

Adults on non-ICU wards were eligible if they had at least one feature of mild acute respiratory failure: new radiological consolidation or atelectasis; oxygen saturation below 92% on room air; an arterial oxygen tension to inspired oxygen fraction ratio below 300 mmHg; mild hypercapnic acidaemia; or clinical respiratory distress on room air. Severe hypoxaemia, defined as a ratio below 200 mmHg, and hypercapnia with pH below 7.30 were exclusions. Other important exclusions included a COPD exacerbation as the cause, immediate need for ventilation or ICU admission, an extremely poor short-term prognosis and contraindications to non-invasive ventilation.

The eligibility definition was consequently broader than a requirement for arterial hypoxaemia in every patient. Radiological changes or distress could qualify, while oxygenation was often assessed using pulse oximetry. Prior invasive or non-invasive ventilation for respiratory failure during the same admission was excluded. Contraindications included inability to protect the airway, unmanageable secretions, vomiting, shock and several recent upper airway or gastrointestinal procedures.⁵

Patients allocated to early treatment received two-hour cycles of non-invasive ventilation every eight hours, in addition to treatment of the underlying cause. CPAP at 5–8 cmH₂O was preferred; bilevel support was used for hypercapnia. Clinicians adjusted settings to target saturation above 92% and arterial carbon dioxide below 45 mmHg. Between sessions, patients breathed room air or received conventional low-flow oxygen. The programme continued for at least 24 hours despite early improvement, unless deterioration, intolerance or another stopping criterion intervened, and for a maximum of 12 days.

Controls received conventional oxygen and other indicated care, including antibiotics or diuretics. Non-invasive ventilation could be started when clinically necessary after the patient reached the primary endpoint. Prolonged rescue treatment meant at least 12 hours daily. ICU physicians helped decide whether deterioration should be managed with prolonged support on the ward or escalation to intensive care. High-flow nasal oxygen was not part of the assigned strategy and its use was classified as a deviation.

The primary endpoint was progression to severe acute respiratory failure during the index hospitalisation. This required severe hypoxaemia, marked persistent respiratory distress while receiving oxygen, or arterial carbon dioxide above 45 mmHg with pH below 7.30. Evaluations occurred three times on the first day and daily thereafter, before treatment cycles in the early-support group. Arterial blood gases were obtained at the clinician's discretion. Patients mistakenly enrolled with severe hypoxaemia were handled differently: the supplement counted their primary endpoint only if respiratory deterioration led to intubation, ICU admission or death for respiratory reasons.

Prespecified secondary outcomes were 28-day all-cause mortality, hospital length of stay and a composite of other respiratory complications. That composite included atelectasis, hospital-acquired pneumonia, pneumothorax, pleural effusion, intubation for respiratory reasons, tracheostomy and acute respiratory distress syndrome. Prolonged non-invasive support, ICU admission and 90-day mortality were presented as exploratory outcomes. Telephone follow-up assessed survival at 28 and 90 days.

The planned sample assumed a reduction in severe respiratory failure from 22% to 11%, with 90% power and a two-sided 5% significance level. A calculated requirement of 512 was rounded to 520 to allow for deviations and follow-up losses. The primary analysis compared proportions without adjustment. Additional analyses included a stepwise logistic model, time-to-event analysis and a per-protocol population. No missing-data imputation was used, and the statistician was masked to allocation.

Of 1,317 patients assessed, 524 were randomised: 266 to early support and 258 to usual care. One intervention patient withdrew permission to use data, leaving 265 and 258 in the principal analysis. Ninety-day survival follow-up was complete in this retained population. Median age was 74 versus 73 years, and approximately 45% were women. Pneumonia accounted for 39% versus 46% of respiratory failure, postoperative causes for about one-fifth and cardiogenic pulmonary oedema for 17% in each group. Around one-fifth had COVID-19.

Most participants were alert and not markedly tachypnoeic: median Glasgow Coma Scale score was 15, and respiratory rates were 21 versus 22 breaths/minute. Approximately 28% and 33% were breathing room air at randomisation. Forty-two patients nevertheless had severe hypoxaemia at enrolment, contrary to the intended eligibility criteria. Most recruitment occurred in Italy, which contributed 455 of the 523 retained participants.^{4,5}

CPAP was the reported modality for 238 intervention patients, with median pressure 7 cmH₂O; 27 were assigned bilevel support, with median expiratory and inspiratory

pressures of 5 and 10 cmH₂O. Oronasal masks predominated. Median time to stopping the assigned programme was three days. Nine patients did not receive early support, and 40 did not tolerate it or refused to continue. At least one cycle was missed in 28 patients and shortened in 22. High-flow oxygen was used in three intervention patients and 17 controls.

Severe respiratory failure occurred in 49/265 patients (18.5%) with early support and 73/258 (28.3%) with usual care: relative risk 0.65, 95% confidence interval 0.48–0.90; absolute difference –9.8 percentage points, 95% confidence interval –17.0 to –2.6; $P=0.008$. Severe hypoxaemia occurred in 33 versus 55, and severe distress in 39 versus 62. Components overlapped. Hypercapnic acidaemia occurred in eight versus five.

There were 30 versus 27 deaths by day 28, relative risk 1.08, 95% confidence interval 0.66–1.77. Median hospital stay was ten versus nine days, with no significant difference. Intubation for respiratory reasons occurred in 13 versus 11, and ICU admission in 22 versus 18. Prolonged non-invasive support was used in 30 versus 60, relative risk 0.49, 95% confidence interval 0.33–0.73. Ninety-day deaths were 41 versus 40, approximately 15.5% in each group.

The per-protocol primary result was similar: 45/237 versus 72/235, relative risk 0.62, 95% confidence interval 0.45–0.86. The adjusted odds ratio was 0.56, 95% confidence interval 0.36–0.86. No persuasive treatment interaction emerged in the prespecified subgroups. Respiratory arrest occurred in eight patients per group, unstable haemodynamics in 26 versus 24, and vomiting in eight versus 14. No pressure sores were reported during early treatment.^{4,5}

Critique

NAVIGATE demonstrates fewer episodes meeting its definition of severe respiratory failure with scheduled early non-invasive support. That is a useful finding, but the endpoint must retain its actual meaning. The trial did not demonstrate that the strategy reduced intubation, ICU admission, hospital stay or mortality. It establishes a reduction in measured deterioration more clearly than an improvement in the overall course of illness.⁴

The open-label design matters particularly because persistent respiratory distress was a major component. Dyspnoea, accessory muscle use and paradoxical breathing are clinically relevant, but their ascertainment involves judgement. Knowledge that a patient is receiving preventive support may influence assessment or the threshold for calling deterioration persistent. Masking the statistician cannot remove bias introduced while

outcomes are identified. The parallel reduction in severe hypoxaemia adds objective support, although arterial sampling was discretionary and inspired oxygen estimates during conventional therapy are imperfect.

Assessment timing provides some protection against simply recording the immediate effect of a treatment session: observations were made before cycles. Nevertheless, the measurement schedule became less frequent after day one, and the trial did not standardise all monitoring. It is therefore difficult to determine whether the intervention prevented sustained deterioration, shortened transient episodes or altered their detection. A future study could use more consistent ascertainment and separate symptom burden from treatment-escalation decisions.

Rescue ventilation in controls is essential to the clinical question. This was a comparison of routine early treatment with treatment when deterioration occurred, not a comparison of receiving ventilation with never receiving it. Rescue support might reasonably prevent an episode of severe distress from progressing to intubation. That possibility explains why a difference in the primary endpoint need not produce a mortality benefit, but it does not make the comparator unfair or erase its clinical relevance. Clinicians are deciding whether exposing many mildly affected patients to early treatment improves on a selective rescue strategy.

The lower frequency of prolonged support is potentially valuable, but it cannot establish a reduction in total workload. Scheduled two-hour sessions three times daily were added for nearly the entire intervention group, while only a minority of controls needed prolonged support. Session preparation, mask fitting, surveillance and treatment breaks all consume staff time. The trial did not report a comprehensive comparison of nursing workload, cumulative ventilation hours or resource costs. A claim that routine early support saves ward or ICU resources would go beyond its findings.

Treatment burden was measurable even without that economic assessment. Fifteen per cent could not tolerate treatment or declined to continue, and missed or shortened sessions were not rare. These observations matter when the intended population includes patients with relatively mild abnormalities who may never deteriorate. The absence of pressure sores is reassuring for that specific complication; it does not mean the intervention was comfortable or without opportunity costs. Patient-reported distress and treatment acceptability would strengthen the assessment of net benefit.

The inclusion of 42 patients with severe hypoxaemia creates a more unusual problem. They had already crossed one of the thresholds the trial aimed to prevent, so the investigators used a different endpoint rule for them. Retaining randomised patients

protects against selective exclusion, but applying different outcome definitions makes the overall primary result less uniform. The similar per-protocol result supports the direction of the finding in a narrower population, although exclusions after randomisation introduce their own selection concerns. The trial should not be described as an entirely clean experiment in patients all starting above an oxygenation ratio of 200.⁵

Recruitment over 11 years also affects interpretation. Randomisation preserves concurrent comparisons within the study, but the aggregate estimate spans changes in ward practice, oxygen devices, escalation thresholds and the COVID-19 pandemic. High-flow oxygen was uncommon overall and treated as a deviation, limiting relevance to services where it is an established alternative. Its greater use in controls may reflect evolving practice or rescue needs. Excluding such patients does not recreate a randomised comparison between modern respiratory-support options.

Heterogeneity is relevant because the intervention was predominantly CPAP and the population included pneumonia, pulmonary oedema and postoperative deterioration. These syndromes have different reasons for responding to positive pressure. The trial was not large enough to establish a reliable treatment effect for each diagnosis. Similar subgroup point estimates or non-significant interaction tests do not prove equal benefit. In particular, the findings should neither replace established syndrome-specific indications nor justify indiscriminate support for every new infiltrate or low oxygen saturation.

The claim of safety must remain bounded by precision and setting. The mortality interval allows clinically important benefit or harm, and only 24 patients underwent intubation for respiratory reasons. Similar observed adverse-event counts cannot exclude uncommon but serious consequences of delayed escalation. Moreover, eight centres had over ten years of experience delivering non-invasive ventilation outside intensive care, nine organised formal training, and all had management protocols. ICU physicians directed prolonged support. These service features are part of the context in which the result was obtained, even though staffing was often stretched.

The practical implication is selective rather than universal adoption. A trained ward team with appropriate monitoring and prompt access to escalation may reasonably consider intermittent early support for suitable patients. NAVIGATE strengthens evidence that this can reduce measured progression, but it does not establish a survival advantage or a substitute for intensive care when deterioration warrants it. The decisive next question is whether the added treatment burden produces benefits that patients and healthcare services value beyond avoiding a physiological endpoint.

Body of Evidence

Plant and colleagues randomised 236 patients with mildly or moderately acidotic COPD exacerbations on general respiratory wards. Treatment failure meeting predefined criteria for intubation occurred in 18/118 with additional non-invasive ventilation and 32/118 with standard therapy; in-hospital deaths were 12 versus 24. This demonstrated that ward delivery can improve important outcomes in a well-defined ventilatory-failure syndrome. NAVIGATE deliberately excluded exacerbations of COPD as the cause of respiratory failure, so its uncertainty should not be used to weaken that separate indication.¹

Squadrone and colleagues studied 40 patients with haematological malignancy and early respiratory abnormalities. CPAP reduced ICU admission for mechanical ventilation from 16/20 to 4/20. Two CPAP patients and 14 controls were intubated. These striking results came from a small, highly selected population and supported further investigation rather than a general ward-wide policy. They helped motivate the preventive approach tested in NAVIGATE.²

Wermke and colleagues subsequently identified respiratory failure in 86 patients during surveillance of an allogeneic stem-cell transplant cohort. Early intermittent non-invasive ventilation did not clearly improve ICU admission, intubation or survival compared with oxygen, although inadequate oxygenation was numerically less common. Rescue switching was permitted. This earlier study already illustrated the possibility that improved gas exchange might not translate into better major outcomes when rescue treatment is available.³

In 2015, Lemiale and colleagues randomised 374 immunocompromised patients with hypoxaemic respiratory failure in French and Belgian ICUs. Day-28 mortality was 24.1% with early non-invasive ventilation and 27.3% with oxygen, an absolute difference of –3.2 percentage points, 95% confidence interval –12.1 to 5.6. No clear improvement emerged in other major outcomes. Its participants were sicker and treated in intensive care, but the result reinforces the need to distinguish physiological plausibility from proven survival benefit.⁶

RECOVERY-RS examined respiratory support during COVID-19. In its CPAP comparison, intubation or death within 30 days occurred in 137/377 patients versus 158/356 receiving conventional oxygen, an absolute difference of approximately –8 percentage points, 95% confidence interval –15 to –1. Adverse events were more frequent with CPAP. This was a different severity and treatment setting, with a primary endpoint directly incorporating intubation and death. It provides evidence for a specific clinical context rather than a direct replication of NAVIGATE's intermittent preventive regimen.⁷

COVID-HIGH was closer to a mild-hypoxaemia question but tested a different device. Among 364 randomised patients with COVID-19 pneumonia, escalation of respiratory support occurred in 55/181 with high-flow oxygen and 70/181 with conventional oxygen: relative risk 0.79, 95% confidence interval 0.59–1.05. The trial did not establish reductions in escalation, ICU admission or length of stay. Its uncertainty should not be interpreted as equivalence, and it cannot determine whether high-flow oxygen is better or worse than NAVIGATE's CPAP schedule.⁸

By the end of 2025, the evidence supported matching respiratory support to the syndrome, severity and capabilities of the treating service. NAVIGATE adds an international randomised comparison in a previously under-studied ward population. Its contribution is the reduction in progression to its defined severe respiratory-failure endpoint; the benefit of routinely treating all eligible patients in terms of survival, invasive ventilation and total care burden remains uncertain.

Should intermittent non-invasive ventilation be started routinely for mild respiratory failure on general wards?

Consider it selectively in experienced services. NAVIGATE reduced progression to severe respiratory failure, but did not demonstrate fewer intubations, ICU admissions or deaths.

References

1. Plant PK, Owen JL, Elliott MW. Early use of non-invasive ventilation for acute exacerbations of chronic obstructive pulmonary disease on general respiratory wards: a multicentre randomised controlled trial. *Lancet*. 2000;355(9219):1931-1935.
2. Squadrone V, Massaia M, Bruno B, Marmont F, Falda M, Bagna C, et al. Early CPAP prevents evolution of acute lung injury in patients with hematologic malignancy. *Intensive Care Med*. 2010;36(10):1666-1674.
3. Wermke M, Schiemanck S, Höffken G, Ehninger G, Bornhäuser M, Illmer T. Respiratory failure in patients undergoing allogeneic hematopoietic SCT--a randomized trial on early non-invasive ventilation based on standard care hematology wards. *Bone Marrow Transplant*. 2012;47(4):574-580.
4. Monti G, Cabrini L, Kotani Y, Brusasco C, Kadralinova A, Giardina G, et al. Early noninvasive ventilation in general wards for acute respiratory failure: an international, multicentre, open-label, randomised trial. *Br J Anaesth*. 2025;134(2):382-391.

5. Monti G, Cabrini L, Kotani Y, Brusasco C, Kadralinova A, Giardina G, et al. Supplementary methods and results to Early noninvasive ventilation in general wards for acute respiratory failure: an international, multicentre, open-label, randomised trial. *Br J Anaesth*. 2025.
6. Lemiale V, Mokart D, Resche-Rigon M, Pène F, Mayaux J, Faucher E, et al. Effect of Noninvasive Ventilation vs Oxygen Therapy on Mortality Among Immunocompromised Patients With Acute Respiratory Failure: A Randomized Clinical Trial. *JAMA*. 2015;314(16):1711-1719.
7. Perkins GD, Ji C, Connolly BA, Couper K, Lall R, Baillie JK, et al. Effect of Noninvasive Respiratory Strategies on Intubation or Mortality Among Patients With Acute Hypoxemic Respiratory Failure and COVID-19: The RECOVERY-RS Randomized Clinical Trial. *JAMA*. 2022;327(6):546-558.
8. Crimi C, Noto A, Madotto F, Ippolito M, Nolasco S, Campisi R, et al. High-flow nasal oxygen versus conventional oxygen therapy in patients with COVID-19 pneumonia and mild hypoxaemia: a randomised controlled trial. *Thorax*. 2023;78(4):354-361.



Gastrointestinal Trials

TIMING

Ke L, Li G, Mao W, Doig G, Chen T, Li C, et al. Early versus delayed catheter drainage for patients with necrotizing pancreatitis and early persistent organ failure (TIMING): a multicenter randomized controlled trial. Intensive Care Med. 2025;51:1431–1441.

Introduction

The management of necrotising pancreatitis requires repeated decisions about whether a collection should be left alone, drained or debrided. Organ failure often begins before a necrotic collection has become encapsulated, and systemic inflammation can resemble infection. This creates a difficult clinical situation: a patient remains critically ill, imaging shows an accessible collection, and drainage appears to offer a tangible intervention even when infection is uncertain.

There are plausible reasons for early drainage. A collection may contain inflammatory mediators and activated enzymes, contribute to intra-abdominal pressure or harbour infection before it becomes obvious. Removing fluid could therefore reduce the physiological burden. Against this must be weighed the risks of introducing infection, causing bleeding or fistulation and committing a patient to repeated procedures before necrosis has clearly demarcated. A less invasive procedure is not necessarily an innocuous procedure.¹

The move from primary open necrosectomy towards step-up treatment established the value of limiting procedural injury. It did not establish that intervention should occur earlier. Similarly, evidence about timing after infected necrosis is diagnosed does not completely answer whether persistent organ failure alone should trigger drainage during the second week of disease.

TIMING addressed that latter question. It compared early percutaneous drainage of acute necrotic collections with a strategy of waiting for suspected or confirmed infection and, where possible, encapsulation. The comparison was between different thresholds for intervening, rather than between two mandatory drainage dates for every patient.

Synopsis

TIMING was an investigator-initiated, open-label randomised trial in five Chinese tertiary teaching hospitals with expertise in severe pancreatitis. Recruitment ran from March 2019 to April 2022, with a pause during the COVID-19 pandemic. The Jiangsu Province Key Research and Development Programme funded the study without involvement in design, data interpretation or reporting. Outcome assessors and the trial statistician were masked to allocation.^{1,2}

Patients aged eighteen to seventy years were screened if they had CT-confirmed acute necrotic collections accessible to percutaneous drainage and organ failure persisting for at least 48 hours that remained unresolved on day seven after symptom onset. During days eight to 21, randomisation required worsening of the original organ failure, new organ failure or organ failure persisting for more than seven days. Respiratory, renal and cardiovascular dysfunction were considered.

Important exclusions included pregnancy, chronic or tumour-related pancreatitis, previous drainage or surgery before admission, cardiopulmonary resuscitation during the episode and substantial pre-existing cardiac, respiratory, renal or hepatic disease. The population therefore had severe acute disease but relatively limited major chronic comorbidity, and excluded older adults above seventy years.

Randomisation used computer-generated variable blocks, stratified by site. Assignments were held in sequentially numbered opaque sealed envelopes, opened after consent and audited at trial close-out. Patients and treating clinicians could not be masked because one strategy required an immediate procedure.

Early-intervention patients underwent CT- or ultrasound-guided percutaneous drainage within 24 hours of randomisation. At least one 12–16 French catheter was placed, with additional drains allowed according to anatomy and clinical judgement. Drain output was reviewed daily. Removal was considered after output remained no greater than 50 mL daily for three days when infection was neither suspected nor confirmed. Endoscopic drainage was not permitted as the initial intervention in this arm.

Standard-care patients did not undergo routine immediate drainage. Intervention was deferred until suspected or confirmed infected pancreatic necrosis, preferably after approximately four weeks when the collection was encapsulated. Earlier intervention remained possible when clinically indicated. Percutaneous or endoscopic drainage could be selected locally, followed by minimally invasive debridement if required. Both groups received organ support, nutrition and antibiotics for suspected or confirmed infection; prophylactic antibiotics were not part of the strategy.

The primary outcome was a composite of death or major complications during the index admission. Major complications comprised new organ failure, bleeding requiring intervention and gastrointestinal perforation or fistula requiring intervention. Secondary outcomes included the individual components, days alive without organ failure to day 21, intra-abdominal pressure, infected necrosis, sepsis, pancreatic fistula, procedure use, costs and 90-day outcomes.

The sample calculation assumed a primary event rate of 50% under standard care and a 25-percentage-point absolute reduction with early intervention, based on a thirty-patient pilot study. A sample of 120 was estimated to provide 80% power. The main analysis retained all randomised patients, with additional adjusted and per-protocol analyses.

Of 531 patients screened, 120 were randomised: 63 early intervention and 57 standard care. Median age was 42 years and 77% were men. Hypertriglyceridaemia caused pancreatitis in 54%. Randomisation occurred a median of nine days after symptom onset. Baseline ventilation was more common in the early group, 70% versus 54%, and about one-third received kidney replacement therapy.

Two early-intervention patients did not undergo drainage, one because the family declined the procedure and one because the patient improved overnight. One standard-care patient underwent early intervention without the specified indication. Follow-up was retained in all three, and all 120 were followed through index discharge or death. Fifty-five participants came from the coordinating hospital and 37 from the next largest centre.

The primary outcome occurred in 21/63 early-intervention patients (33.3%) and 21/57 standard-care patients (36.8%), a risk difference of -3.5 percentage points (95% CI -20.6 to 13.6; $P=0.69$). Adjustment for prespecified prognostic variables did not establish benefit. In-hospital mortality was 14/63 (22.2%) versus 11/57 (19.3%), and 90-day mortality was fourteen versus twelve deaths.

Median days alive without organ failure to day 21 were four versus one, but the distributions did not differ clearly. There was no demonstrated improvement in intra-abdominal pressure, ICU- or hospital-free days, infected necrosis, sepsis or pancreatic fistula. Collection dimensions decreased after early drainage, showing that the intervention removed fluid, without establishing a corresponding patient-centred benefit.

Sixty-one of 63 early-intervention patients underwent an invasive intervention, compared with 25 of 57 under standard care. Thirty-two standard-care patients (56.1%) avoided invasive intervention entirely. Among those who underwent a procedure, median time from disease onset to the first intervention was nine versus seventeen days. Minimally invasive debridement was required in eighteen versus twelve patients and open surgery in six versus four, without clear differences.

Adverse events were reported in sixteen early-intervention patients (25.4%) and seven controls (12.3%); serious events occurred in three and one, respectively. The published significance claim for overall adverse events is inconsistent with the displayed counts and stated test, so the numerical imbalance requires caution rather than being treated as established statistical evidence of harm.^{1,2}

Critique

TIMING does not support routine early drainage solely because organ failure persists in a patient with an accessible acute necrotic collection. It achieved a clear difference in procedural exposure without demonstrating a reduction in major complications, death or organ failure duration. The ability of more than half the standard-care group to avoid intervention is especially relevant when considering whether a visible collection necessarily requires treatment.

The trial is nevertheless too small to establish equivalence or exclude meaningful benefit or harm. The primary interval ranges from approximately 21 fewer to fourteen more events per hundred patients. The mortality interval is similarly wide. The correct conclusion is failure to demonstrate the proposed advantage, rather than proof that timing can never matter or that every early procedure causes harm.

The sample-size assumption was particularly ambitious. Halving a composite event rate from 50% to 25% was based on a pilot with only thirty participants. Such studies are useful for assessing feasibility, but their effect estimates are unstable and can exaggerate the benefit eventually observed. The actual control event rate was also lower than expected. The completed trial therefore had limited ability to detect smaller effects that clinicians could still consider worthwhile.

The treatment comparison is best described as routine early drainage versus selective deferred intervention. Standard care did not require every patient to wait exactly four weeks, and the patients who eventually underwent a procedure did so at a median of seventeen days. Conversely, 56% never required one. A headline of “day nine versus day

seventeen” omits those who avoided the procedure and understates the difference between the policies.

The timing medians are calculated only among patients who received procedures. Their composition depends on allocation and subsequent illness, so the eight-day difference is not a randomised estimate of the effect of delaying the same intervention in otherwise identical patients. In the early arm, drainage was almost universal; in standard care, it identified patients with a later clinical indication. The trial's primary allocation-based comparison remains the appropriate basis for judging the overall strategy.

This also explains why the possibility of avoiding drainage is a treatment benefit in its own right, but not proof of improved survival. Some collections resolve while organ support and treatment of the underlying disease continue. Routine intervention exposes those patients to a procedure they might never need. The trial did not demonstrate lower total costs or fewer later debridements with early drainage, so removal of fluid should not be assumed to simplify the subsequent clinical course.

Open-label treatment creates opportunities for differences in co-intervention. Decisions about antibiotics, repeat drainage, necrosectomy and discharge were made by clinicians aware of allocation. Blinded outcome assessment reduces bias in classifying endpoints, but cannot undo treatment differences created before assessment. The composite also includes bleeding and fistula only when intervention was required, linking some outcomes to clinical thresholds as well as biological events.

Infection ascertainment is particularly difficult. Early drainage provides material for culture in almost every intervention patient, while a patient managed conservatively may never have a comparable sample. Organisms recovered from a drain can also become harder to interpret over time. The numerical infected-necrosis difference, 50.8% versus 42.1%, could reflect several mechanisms, including true infection, differential detection or chance. It does not establish that early drainage introduces infection, nor prove that it does not.

The main hypothesis concerned reducing systemic inflammation or pressure by evacuating a collection. Drainage volumes and smaller subsequent collection dimensions confirm anatomical activity, but no clear improvement in intra-abdominal pressure or organ failure followed. This weakens the assumption that reducing visible fluid necessarily removes the dominant cause of ongoing critical illness. Catheters may also drain liquid without removing substantial solid necrosis or interrupting the wider inflammatory process.

The primary composite deserves separate consideration. Death, new organ failure, bleeding and fistula differ in importance and mechanism. Components can occur together, and a reduction in one can coexist with an increase in another. Here, the overall estimate was slightly favourable while mortality was numerically higher with early intervention, both with broad intervals. No individual component established a benefit. The composite should therefore not be translated into a survival claim.

New organ failure may also be an imperfect measure of recovery in a population selected for existing persistent failure. A treatment that shortens established respiratory failure without preventing a new failing organ might not strongly affect that component. The secondary days-alive-without-organ-failure measure addresses this concern more directly, but was also inconclusive. Its median difference alone, four versus one days, does not demonstrate a three-day treatment benefit when the full distributions and test do not support that inference.

Baseline imbalance is unsurprising in a small trial. More early-intervention patients were ventilated, and their collections had somewhat greater long-axis dimensions. Prespecified adjustment and an additional ventilation-adjusted analysis did not establish benefit. Adjustment cannot ensure balance in every measured or unmeasured feature, and selecting covariates because of baseline P values is less satisfactory than a fully prespecified prognostic model. These analyses are supportive checks, not a means to extract a definitive result from limited information.

The adverse-event reporting discrepancy is material. Both the main text and supplementary table give sixteen of 63 versus seven of 57 patients and report $P=0.007$, with Fisher's exact test specified. A two-sided Fisher calculation from those counts gives approximately $P=0.103$. The displayed data therefore do not reproduce the published significance claim. The reason cannot be resolved from the supplied report; the counts can be described as a numerical excess, while any assertion of proven harm should await clarification.

That discrepancy does not imply that early drainage is safe. Serious events were uncommon, their comparison was imprecise, and individual adverse events may have varied in attribution. It does mean that the case against routine early intervention should rest primarily on lack of demonstrated benefit and substantially greater procedure exposure, rather than on an unreconciled P value.

Generalisability is limited by case mix and setting. Participants were young, predominantly male and frequently had hypertriglyceridaemic pancreatitis. Older patients and those with major chronic disease were excluded. Two hospitals recruited

most of the cohort, and all sites had relevant expertise. Different aetiologies, frailty, anatomy and procedural resources could alter both the chance of spontaneous recovery and the risks of drainage. The small subgroup analyses cannot rule out such differences.

The trial also does not test every approach to early intervention. Initial endoscopic drainage was prohibited in the early arm, and drainage of a liquid component is different from early necrosectomy. Results should not be extrapolated to mandatory waiting in a patient with an urgent complication or uncontrolled infection requiring source control. The tested policy retained clinically indicated intervention; it was not passive observation regardless of deterioration.

TIMING consequently supports a selective strategy while leaving substantial uncertainty about narrower high-risk phenotypes. Persistent organ failure is an important prognostic marker, but is not by itself evidence that an acute necrotic collection is the reversible cause. The decision to intervene still depends on the evolving clinical indication, anatomy and multidisciplinary assessment, rather than the presence of organ failure plus a technically accessible target alone.

Body of Evidence

PANTER established the value of a minimally invasive step-up approach compared with primary open necrosectomy in infected necrotising pancreatitis. Among 88 patients, major complications or death occurred in 40% with step-up treatment and 69% with open surgery, with a risk ratio of 0.57 (95% CI 0.38–0.87). Thirty-five per cent of step-up patients were managed with drainage alone. Mortality itself did not differ clearly. This trial changed the intensity and sequence of intervention, rather than establishing an indication for routine early drainage of potentially sterile collections.³

TENSION compared endoscopic with surgical step-up treatment in 98 patients with infected necrosis and an indication for intervention. Major complications or death occurred in 43% versus 45%, with no superiority for the primary composite. Endoscopic treatment produced fewer pancreatic fistulas and a shorter hospital stay. The study shows that technical route can influence some outcomes even when the broad composite is similar; it does not prove routes are interchangeable in TIMING's earlier, often unencapsulated collections.⁴

The thirty-patient pilot preceding TIMING offered a hypothesis rather than confirmation. Major complications or death occurred in three of fifteen patients assigned early on-demand drainage and seven of fifteen controls, a relative risk of 0.43 (95% CI 0.14–1.35). Only eight and four patients, respectively, actually underwent early drainage.

The wide interval and non-significant result were consistent with anything from substantial benefit to no useful effect. Using its large point estimate to power the subsequent trial left that study vulnerable to imprecision when the apparent effect became smaller.⁵

POINTER addressed timing after infected necrosis had been diagnosed. It randomised 104 patients to immediate drainage or postponement until walled-off necrosis when possible. The six-month Comprehensive Complication Index was similar, with mean scores of 57 and 58 and a difference of -1 (95% CI -12 to 10). Immediate drainage led to more interventions, averaging 4.4 versus 2.6. Nineteen of 49 postponed-treatment patients never required drainage, and seventeen of those survived.⁶

POINTER and TIMING are complementary rather than duplicate trials. POINTER required infected necrosis and generally considered intervention later in the disease course; TIMING selected patients for early persistent organ failure, with drainage around day nine regardless of a mandatory infection diagnosis. Both failed to show that automatically intervening earlier improved their primary outcomes, and both demonstrated that selective delay could avoid procedures in some patients. Neither establishes that source control should be withheld when a compelling urgent indication exists.

The broader evidence therefore supports separating three questions: whether intervention is required, which technique should be used and when it should occur. PANTER and TENSION mainly inform technique and escalation once an indication exists. POINTER and TIMING inform thresholds and timing in different clinical populations. Combining them into a blanket rule about all pancreatic collections would lose the distinctions that make the trials useful.

Future studies need more realistic effect-size assumptions, adequate representation of older and more varied patients, and clear separation of sterile inflammatory failure from infection-driven deterioration. Complete reporting of cumulative procedural burden and long-term pancreatic function would help determine whether an early anatomical intervention changes recovery rather than simply moving procedures forward. Until such evidence emerges, TIMING provides no reason to make persistent organ failure alone a routine trigger for early catheter drainage.

Should persistent early organ failure routinely trigger drainage of an acute necrotic pancreatic collection?

TIMING does not support routine early drainage on this basis. Selective deferred intervention avoided procedures in many patients without a demonstrated disadvantage in the primary outcome, while urgent clinical indications still require individual assessment.

References

1. Ke L, Li G, Mao W, Doig G, Chen T, Li C, et al. Early versus delayed catheter drainage for patients with necrotizing pancreatitis and early persistent organ failure (TIMING): a multicenter randomized controlled trial. *Intensive Care Med.* 2025;51(8):1431-1441.
2. Ke L, Li G, Mao W, Doig G, Chen T, Li C, et al. TIMING supplementary appendix. *Intensive Care Med.* 2025. doi:10.1007/s00134-025-08020-x.
3. van Santvoort HC, Besselink MG, Bakker OJ, Hofker HS, Boermeester MA, Dejong CH, et al. A step-up approach or open necrosectomy for necrotizing pancreatitis. *N Engl J Med.* 2010;362(16):1491-1502.
4. van Brunschot S, van Grinsven J, van Santvoort HC, Bakker OJ, Besselink MG, Boermeester MA, et al. Endoscopic or surgical step-up approach for infected necrotising pancreatitis: a multicentre randomised trial. *Lancet.* 2018;391(10115):51-58.
5. Ke L, Dong X, Chen T, Doig GS, Li G, Ye B, et al. Early on-demand drainage or standard management for acute pancreatitis patients with acute necrotic collections and persistent organ failure: A pilot randomized controlled trial. *J Hepatobiliary Pancreat Sci.* 2021;28(4):387-396.
6. Boxhoorn L, van Dijk SM, van Grinsven J, Verdonk RC, Boermeester MA, Bollen TL, et al. Immediate versus Postponed Intervention for Infected Necrotizing Pancreatitis. *N Engl J Med.* 2021;385(15):1372-1381.

Empirical antifungals in ACLF

Verma N, Valsan A, Garg P, Sarabu S, Kaur P, Mohan N, et al. Empirical Antifungal Therapy Improves Survival in Patients With Acute-on-Chronic Liver Failure With Suspected Invasive Fungal Infections: A Pragmatic Randomized Trial. Am J Gastroenterol. 2026;121(9):2198–2210.

Introduction

Suspected invasive fungal infection presents a particularly difficult treatment decision in acute-on-chronic liver failure (ACLF). Organ dysfunction, recent antibiotics, invasive devices and impaired host defences may increase the likelihood and consequences of infection. Meanwhile, fever, circulatory deterioration and pulmonary infiltrates have numerous competing explanations. Fungal cultures may be slow or negative, and a positive biomarker does not necessarily establish that invasive fungal disease is causing the deterioration. Waiting for greater diagnostic certainty may lose a treatment opportunity; treating everyone exposes many patients to unnecessary therapy.

The relevant choice is therefore a treatment strategy, not simply an antifungal drug. Immediate empirical treatment based on a combination of risk factors and persistent sepsis can be compared with treatment triggered by additional laboratory, radiological or mycological evidence. Both approaches require active investigation and subsequent reassessment. Starting early and stopping when evidence fails to support infection is a different policy from indiscriminate prolonged antifungal use.

Previous intensive care trials have not established a general survival benefit from empirical antifungals or biomarker-triggered earlier treatment. However, their populations differ substantially from patients with severe ACLF at a centre with a high burden of fungal disease. An earlier ACLF cohort demonstrated frequent invasive fungal infection and an association with poor survival, but could not establish whether an earlier treatment threshold would improve outcomes.^{1,2,3}

Verma and colleagues compared suspicion-based empirical antifungal treatment with a diagnostic or biomarker-driven pre-emptive strategy in this specific population. The principal report first appeared online in November 2025, despite its 2026 print citation.⁴

Synopsis

This was a pragmatic, single-centre, parallel-group randomised trial at the Postgraduate Institute of Medical Education and Research in Chandigarh, India. Recruitment ran from November 2019 to December 2024. The Indian Council of Medical Research supported the study. Treatment was open label, with blinded endpoint adjudication and separate assessment of adverse drug reactions. Patients were allocated equally using a computer-generated sequence and opaque envelopes managed by an independent pharmacologist.⁴

Eligible patients were adults with ACLF who had been in intensive care for more than 48 hours or had relevant preceding hospital exposure, at least two specified fungal host risk factors, and persistent or worsening clinical features raising suspicion of invasive fungal infection. These features included unresolved sepsis or systemic inflammatory response despite more than three days of broad-spectrum antibiotics. Risk factors included prolonged ventilation, vascular catheters, renal replacement therapy, diabetes, parenteral nutrition and advanced liver disease. Exclusions included neutropenia, recent antifungal treatment, active malignancy, HIV infection and very advanced illness considered unlikely to permit meaningful treatment, including four or more organ failures.^{4,5}

The empirical group was to start antifungal treatment at enrolment. The pre-emptive group started when additional diagnostic criteria were met. These included serum β -D-glucan or galactomannan thresholds, compatible imaging, mycological findings or specified clinical evidence. A single β -D-glucan concentration of at least 150 pg/mL, or repeated values of at least 80 pg/mL, could trigger treatment; corresponding galactomannan index thresholds were 1.0 for a single result or 0.5 for serial results. The comparator therefore did not require a positive sterile-site culture or histological proof. It was a relatively broad investigation-driven strategy.^{4,5}

Liposomal amphotericin B, usually 3–5 mg/kg daily, was the predominant initial treatment. Other agents could be selected or substituted according to clinical and microbiological circumstances. Drug choice was not randomised. Both groups underwent ongoing clinical, biomarker and culture assessment, with opportunities to stop or de-escalate treatment when serial investigations did not support invasive infection. This stewardship component is important to the intervention's definition.^{4,5}

The primary endpoint was overall survival at 28 days in the intention-to-treat population. Kaplan–Meier curves and a log-rank comparison were used, with Cox regression to estimate the mortality hazard ratio and examine adjustment for liver disease severity. The report describes censoring patients who were alive, discharged or transplanted at

the corresponding time point. Secondary outcomes included in-hospital and fungal infection-attributable mortality, severity scores, treatment success, organ support, adverse events and cost effectiveness. Treatment success combined survival for at least seven days with resolution of infection, absence of breakthrough infection and no treatment-limiting toxicity.^{4,5}

The sample size assumed 28-day survival of 46% with empirical treatment and 27% with pre-emptive treatment. A calculation of 198 patients, with 80% power and a two-sided significance level of 0.05, was increased to 216 to allow for attrition. All 216 randomised patients, 108 in each group, entered the primary analysis. The flow diagram records no loss to follow-up. No patient underwent liver transplantation.⁴

Patients were relatively young, with median ages of approximately 42 and 43 years, and about 92% were men. Alcohol-related cirrhosis predominated. Median Model for End-Stage Liver Disease–sodium scores were approximately 31, and median organ failure scores were 12. Most met European Association for the Study of the Liver ACLF criteria; a minority met the alternative Asian definition without European ACLF. All had unresolved sepsis or systemic inflammation. Approximately a quarter were mechanically ventilated at baseline. This was a selected population with advanced liver disease, rather than an unselected cohort of patients receiving antifungals in intensive care.⁴

Antifungal treatment was received by 107/108 empirical-group patients and 89/108 pre-emptive-group patients. Among the 19 untreated comparator patients, the flow diagram identifies nine without qualifying biomarkers, five who died before treatment began and five with consent or logistical constraints. Treatment began a median of zero versus three days after randomisation, or four versus 7.5 days after hospital admission. Median treatment duration was eight days in both groups. Liposomal amphotericin B was used in 96 and 80 patients, respectively. De-escalation occurred in 23 and 19.^{4,5}

At 28 days, 38/108 patients assigned empirical treatment were alive, compared with 14/108 assigned pre-emptive treatment: 35.2% versus 13.0%. Corresponding deaths were 70/108 and 94/108. The mortality hazard ratio was 0.64 (95% confidence interval [CI] 0.47–0.88), with a log-rank P value of 0.005. The absolute difference in survival was approximately 22 percentage points. Adjustment separately for major severity measures produced similar hazard ratios.⁴

The proportional hazards assumption was violated, with a global Schoenfeld test P value of 0.001. The investigators therefore also reported restricted mean survival time through day 28. The estimated difference was 2.5 additional days with empirical

treatment (95% CI 0.04–5.0; $P=0.046$). This describes average survival time gained within the observation window and should not be confused with survival beyond 28 days.⁴

In-hospital deaths were 60/108 versus 77/108, while deaths attributed to invasive fungal infection were 60/108 versus 82/108. These are different endpoints, despite their conflation in the abstract. Treatment success was recorded in 40/107 treated empirical-group patients and 15/89 treated comparator patients. Median ventilator-free days were 8.5 versus 7.0, without a statistically clear difference; neither intensive care nor hospital length of stay differed clearly.^{4,5}

Adverse events affected 98/108 versus 107/108 patients. Serious events included death and septic shock, so this overall comparison partly reflects disease outcomes. Acute kidney injury occurred in 27 versus 26 patients and hypokalaemia in 26 versus 25. Exploratory analyses examined treatment heterogeneity and an economic evaluation suggested favourable cost effectiveness, but both require substantial qualification.^{4,5}

Critique

The trial addresses an important uncertainty in a population poorly represented in larger antifungal trials. Randomisation, equal group sizes, retention of all allocated patients in the primary analysis and an objective primary outcome are meaningful strengths. The treatment contrast was substantial: starting immediately rather than a median of three days later, with a further difference in whether treatment was received at all. A 24-patient difference in 28-day survivors among 216 patients is clinically important and should not be dismissed solely because the study was small or conducted at one centre.

However, the causal contrast needs to be stated accurately. This was a comparison of policies for starting antifungals in highly selected patients with suspected infection. It was not a placebo-controlled test of liposomal amphotericin B, a comparison between antifungal agents, or a trial of treatment in microbiologically proven infection. More than 80% of comparator patients eventually received an antifungal. The result therefore primarily supports a lower threshold and earlier initiation in this setting, rather than demonstrating that all high-risk intensive care patients need antifungal treatment.

The five comparator deaths before treatment began are part of the randomised strategy outcome, not a reason to remove those patients. Excluding them would preferentially retain patients who survived long enough to receive delayed treatment. Nevertheless, their clinical trajectories matter for understanding the mechanism: whether delay reflected negative tests, laboratory turnaround, progressive disease or inability to implement a treatment decision. The aggregate results cannot distinguish these

possibilities completely. Local diagnostic capacity may consequently be a major determinant of how much benefit an empirical policy could deliver elsewhere.

The exceptionally high comparator mortality limits transportability. These were predominantly young men with alcohol-related liver disease, severe liver dysfunction and continuing sepsis despite antibiotics. Transplantation did not occur. Patients with several alternative immunocompromising conditions and the most advanced multiorgan failure were excluded. The observed absolute benefit should not be carried directly into a transplant centre with earlier access to diagnosis, a different fungal ecology or a population with less severe hepatic disease. Conversely, this extreme baseline risk may explain why a signal emerged here when broader intensive care trials were neutral.

Diagnostic classification adds uncertainty. Most patients were classified as having probable rather than proven fungal infection, and the categories were assigned after randomisation. These classifications cannot establish that both groups had equivalent invasive infection at baseline. Biomarker positivity, non-specific imaging and cultures from non-sterile sites have different evidential weight from bloodstream infection or tissue invasion. The study's clinical question remains valid under diagnostic uncertainty, but its result should not be presented as a treatment effect exclusively in confirmed fungal disease.^{4,5}

Open-label management introduces potential differences in co-interventions, repeat testing and decisions to continue or stop treatment. Blinded endpoint adjudication helps with classification, particularly attribution of death, but cannot erase treatment-dependent care. All-cause survival is therefore more persuasive than fungal infection-attributable mortality. The latter requires a causal judgement in patients with liver failure, bacterial infection and overlapping organ dysfunction, where several processes may contribute to death.

The survival analysis also warrants clarification. The primary text states that patients were followed for 28 days, and the flow diagram records complete follow-up, but both the methods and supplementary endpoint definition describe discharge as a censoring event. There were more 28-day deaths than in-hospital deaths, suggesting that at least some post-discharge outcome information was incorporated. The exact handling of discharge and subsequent deaths is not sufficiently clear. If discharge ends observation, independent censoring is a substantive assumption because discharge is strongly related to prognosis. This ambiguity should be resolved before treating the Kaplan–Meier estimate as an unequivocally complete assessment of all-cause 28-day survival.^{4,5}

Violation of proportional hazards does not invalidate the randomised comparison, but it changes interpretation of the hazard ratio. A single value of 0.64 should not be described as a constant reduction in the instantaneous risk of death throughout follow-up. The restricted mean survival estimate is a useful complementary summary and shows a less dramatic-looking, though potentially important, gain of 2.5 days within four weeks. Its confidence interval is wide and barely excludes zero. The two summaries answer different questions; neither should be selected solely because it sounds more favourable.

Several secondary analyses are less secure. Treatment success used denominators restricted to patients who actually received therapy, which removes the protection of the original allocation and excludes proportionally more comparator patients. The lower overall adverse-event count includes deaths and septic shock and is not proof that earlier antifungal exposure is intrinsically less toxic. Similar kidney injury and hypokalaemia counts provide more direct, though imprecise, information about relevant harms. Resistance, ecological consequences and longer-term toxicity were not established.

The economic claim cannot presently be accepted at face value. The supplementary analysis specifies a 28-day horizon, yet gives total quality-adjusted life-years of 29.9 and 10.1 in groups of 108 patients. Even if every patient survived all 28 days in perfect health, a group could accrue only approximately 8.3 quality-adjusted life-years within that horizon. The reported totals therefore require an explanation of their units, time horizon or calculation. This is not a marginal issue in the interpretation of an incremental cost-effectiveness ratio: its denominator determines the result. Higher survival may justify additional expenditure, but this particular economic calculation does not reliably quantify that judgement.⁵

Finally, the exploratory causal tree is unsuitable for bedside treatment selection. It divides a small trial into still smaller groups, with very large apparent benefits in some terminal branches and no external validation. Inconsistencies in the supplementary coding descriptions further complicate interpretation. Conventional subgroup analyses did not provide convincing evidence of a reproducible treatment interaction. Neither the tree nor comparisons between survivors and non-survivors establish that a particular age or organ-failure pattern should determine treatment. In particular, longer treatment among survivors may simply reflect having survived long enough to receive it.

Body of Evidence

Schuster and colleagues randomised 270 intensive care patients with persistent fever despite broad-spectrum antibiotics, central venous access and substantial illness severity to fluconazole or placebo. Among 249 evaluable patients, the composite treatment-success outcome occurred in 44/122 versus 48/127, relative risk 0.95 (95% CI 0.69–1.32). Documented invasive candidiasis was numerically less frequent with fluconazole, but uncertainty was considerable. This trial demonstrated the difficulty of identifying a responsive population from common intensive care risk factors and persistent fever alone. It did not specifically address severe ACLF or a predominantly amphotericin-based strategy.⁶

The INTENSE trial tested pre-emptive micafungin after emergency gastrointestinal surgery for intra-abdominal infection. The full analysis set contained 241 patients. Invasive candidiasis occurred in 11.1% with micafungin and 8.9% with placebo, an absolute difference of 2.24 percentage points (95% CI –5.52 to 10.20). The trial did not establish prevention of invasive candidiasis. Its surgical population and prevention-oriented question differ from treatment of persistent sepsis in ACLF, but it reinforces that a biologically plausible high-risk group does not guarantee clinical benefit from earlier antifungals.⁷

EMPIRICUS provides a particularly important comparison. It enrolled 260 non-neutropenic, non-transplanted patients with intensive care-acquired sepsis, multiple *Candida* colonisation and multiorgan failure across 19 French units. Patients received empirical micafungin or placebo. In the modified intention-to-treat population, 28-day survival without proven invasive fungal infection was not clearly improved: the hazard ratio was 1.35 (95% CI 0.87–2.08), favouring micafungin for this survival endpoint. New invasive fungal infections were reduced from 15/123 to 4/128, but prevention of infection did not establish improvement in the primary patient outcome.¹

This distinction matters when interpreting the ACLF trial. EMPIRICUS selected patients using colonisation and organ dysfunction, whereas Verma and colleagues studied advanced hepatic disease with persistent sepsis in a high-burden setting. The antifungal spectrum, diagnostic criteria and background risks also differed. EMPIRICUS therefore argues against universal empirical treatment, but cannot exclude a benefit in the more specific ACLF population. Equally, the positive single-centre ACLF result cannot overturn the neutral multicentre evidence for the broader intensive care population.

CandiSep examined whether β -D-glucan-guided treatment could improve the timing of antifungals in adults with sepsis at risk of invasive *Candida* infection. Across 18 intensive care units, 339 patients were randomised to biomarker-guided or culture-driven

treatment. Antifungals were started earlier, at a median of 1.1 versus 4.4 days, and many more patients received them within 96 hours. Nevertheless, 28-day mortality was 58/172 versus 51/167, relative risk 1.10 (95% CI 0.80–1.51). Earlier exposure alone did not improve survival in that population.²

The earlier retrospective ACLF study is more directly relevant to disease selection. Among 264 patients, invasive fungal infection was diagnosed in 39, including both candidiasis and aspergillosis. Prior antibiotics and haemodialysis were associated with infection, and fungal infection was associated with mortality. β -D-glucan was sensitive but had limited specificity at the studied threshold. These findings supported the need for a randomised strategy trial while illustrating why positive biomarkers do not eliminate diagnostic uncertainty. The cohort supplies rationale, not independent confirmation of a treatment benefit.³

Taken together, the evidence supports a cautious, population-specific interpretation. The new trial offers a substantial survival signal for immediate empirical therapy in severely ill patients with ACLF and persistent sepsis at a high-burden centre. It does not justify routine empirical antifungals across intensive care, determine the best agent, validate a subgroup treatment rule or establish cost effectiveness. Replication with clearly documented post-discharge follow-up, rapid diagnostic pathways, contemporary stewardship and diverse centres would materially strengthen the case for broader implementation.

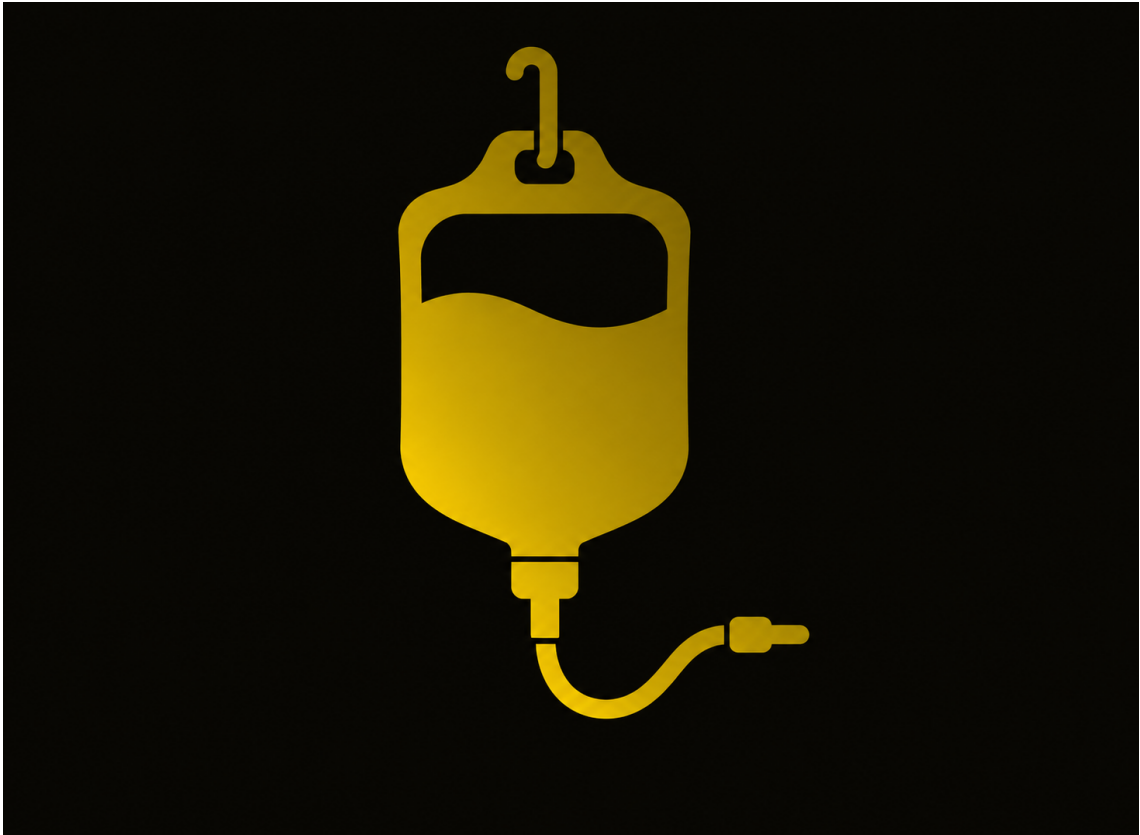
Should empirical antifungals start immediately in ACLF with persistent sepsis and suspected fungal infection?

This trial supports considering earlier treatment in carefully selected patients resembling its high-risk population, alongside active diagnostics and stewardship. It is insufficient to mandate routine empirical therapy across centres or validate its proposed economic and subgroup claims.

References

1. Timsit JF, Azoulay E, Schwebel C, Charles PE, Cornet M, Souweine B, et al. Empirical Micafungin Treatment and Survival Without Invasive Fungal Infection in Adults With ICU-Acquired Sepsis, Candida Colonization, and Multiple Organ Failure: The EMPIRICUS Randomized Clinical Trial. *JAMA*. 2016;316(15):1555-1564.

2. Bloos F, Held J, Kluge S, Simon P, Kogelmann K, de Heer G, et al. (13)- β -D-Glucan-guided antifungal therapy in adults with sepsis: the CandiSep randomized clinical trial. *Intensive Care Med.* 2022;48(7):865-875.
3. Verma N, Singh S, Taneja S, Duseja A, Singh V, Dhiman RK, et al. Invasive fungal infections amongst patients with acute-on-chronic liver failure at high risk for fungal infections. *Liver Int.* 2019;39(3):503-513.
4. Verma N, Valsan A, Garg P, Sarabu S, Kaur P, Mohan N, et al. Empirical Antifungal Therapy Improves Survival in Patients With Acute-on-Chronic Liver Failure With Suspected Invasive Fungal Infections: A Pragmatic Randomized Trial. *Am J Gastroenterol.* 2026;121(9):2198-2210.
5. Verma N, Valsan A, Garg P, Sarabu S, Kaur P, Mohan N, et al. Supplementary material to: Empirical Antifungal Therapy Improves Survival in Patients With Acute-on-Chronic Liver Failure With Suspected Invasive Fungal Infections: A Pragmatic Randomized Trial. *Am J Gastroenterol.* 2026;121(9):2198–2210.
6. Schuster MG, Edwards JE Jr, Sobel JD, Darouiche RO, Karchmer AW, Hadley S, et al. Empirical fluconazole versus placebo for intensive care unit patients: a randomized trial. *Ann Intern Med.* 2008;149(2):83-90.
7. Knitsch W, Vincent JL, Utzolino S, François B, Dinya T, Dimopoulos G, et al. A randomized, placebo-controlled trial of preemptive antifungal therapy for the prevention of invasive candidiasis following gastrointestinal surgery for intra-abdominal infections. *Clin Infect Dis.* 2015;61(11):1671-1678.



Nutrition Trials

TARGET Protein

Summers MJ, Chapple LS, Karahalios A, Bellomo R, Chapman MJ, Ferrie S, et al. Augmented Enteral Protein During Critical Illness: The TARGET Protein Randomized Clinical Trial. JAMA. 2025;334(4):319-328.

Introduction

Loss of muscle during critical illness provides a compelling rationale for supplying more protein. The practical question is whether additional enteral protein can alter that loss and improve recovery while acute inflammation, organ dysfunction and immobility continue. A patient may receive more amino acids without incorporating them into functional muscle. Increased delivery could instead increase nitrogen disposal and metabolic work. The distinction between nutritional adequacy and a beneficial treatment effect therefore requires a clinical comparison.

Protein is also difficult to study separately from energy. Increasing the volume of an ordinary feed increases both, while additional supplements introduce changes in administration and adherence. Two commercially available, isocaloric formulae with different protein concentrations offer a relatively straightforward way to increase protein without deliberately increasing calories. A preceding blinded feasibility trial showed that this approach could produce a substantial difference in protein delivery. It did not establish that the additional protein improved patient outcomes.¹

The uncertainty became more important as larger trials appeared. EFFORT Protein found no improvement in live hospital discharge with a higher prescription in nutritionally high-risk patients; PRECISE subsequently found worse health-related quality of life with higher enteral protein. These studies challenged the assumption that closing a protein deficit necessarily helps, while leaving questions about dose, patient selection and the phase of illness.^{2,3}

TARGET Protein tested a policy that could be implemented through the feed routinely stocked in an intensive care unit. Rather than selecting patients expected to require prolonged ventilation or individually prescribing a high weight-based dose, it enrolled a broad population starting enteral nutrition. The trial asked whether this pragmatic increase in protein concentration would increase days alive and free of the index hospital by day 90.⁴

Synopsis

TARGET Protein was an investigator-initiated, open-label, cluster-randomised crossover trial in eight Australian and New Zealand intensive care units. Recruitment ran from 23 May 2022 to 23 August 2023. Four units started three months before the others, and each recruited for twelve months. Public and charitable funding was supplemented by an unrestricted Nutricia grant supporting formula purchases and dissemination. The company received the protocol before commencement and the report before submission; the investigators stated that funders had no role in trial conduct, analysis or the publication decision.⁴

Patients aged at least sixteen years were eligible when commencing enteral nutrition during an intensive care admission, including first initiation during a subsequent admission within the same hospitalisation. Exclusions were a clinical contraindication to the trial formula, more than twelve hours of non-trial enteral feeding, or previous participation. There was no requirement for invasive ventilation or a predicted minimum feeding duration. Australian sites used a waiver of consent; the New Zealand site used initial waiver arrangements with subsequent consent to continue where applicable.^{4,5}

A computer-generated schedule allocated units to alternating three-month periods of augmented or usual protein. Four units began with each formula, with allocation balanced within the two staggered starting groups. The sequence was disclosed a month before commencement to permit supply arrangements. A patient continued the originally assigned formula across a unit crossover and, if feeding restarted during a qualifying intensive care readmission, resumed that original assignment. Neither clinicians nor patients were masked to the formula policy.

The augmented formula contained 100 g protein/L and the usual formula 63 g/L. They were isocaloric. Clinicians decided when to start feeding, the target rate and other nutritional management. Sites received education about ideal body weight in patients with overweight or obesity; the supplement recommended a maximum rate of 1 mL/kg ideal body weight/hour. The comparison was therefore between formula policies within usual feeding practice, rather than guaranteed achievement of two prescribed protein doses. Trial feeding continued while indicated, until intensive care discharge, death or ninety days. Clinically necessary alternative feeds were permitted and recorded.

The primary outcome was days free of the index hospital and alive at day 90 after starting trial formula. Days in the initial admission and any readmissions to that hospital were subtracted from ninety; patients dying before day 90 received zero. Death after discharge was identified through health records and national death-index linkage. Admissions to other hospitals were not captured in this outcome. Secondary outcomes

included hospital-free days among survivors, survival, ventilation duration, live intensive care and hospital discharge, tracheostomy, new kidney replacement therapy and discharge destination.

Planning used a cluster crossover design and information from the earlier TARGET energy trial. The original calculation described eight clusters and four periods, with at least sixty patients per cluster-period providing 80% power to detect a one-day difference, and eighty providing 90% power. After completion, the investigators identified a mismatch between how clusters had been specified and the four units assigned to each sequence. They reported that the revised calculation supported greater than 80% power for a two-day difference. This revision did not alter the observed primary estimate.

The primary analysis used individual patient data in a quantile mixed-effects model, with treatment, period and staggered start as fixed effects and intensive care unit as a random effect. The treatment contrast was a difference in medians with a block-bootstrap confidence interval and a two-sided significance threshold of 0.05. Linear and Bayesian models provided secondary analyses. Sensitivity analyses excluded known non-trial formula exposure or admission for palliation or organ donation. An independent committee reviewed safety after two periods at all sites; there was no interim efficacy analysis.^{4,5}

Of 3,599 patients assessed, 3,411 entered the trial. Fourteen withdrew permission to retain data, leaving 3,397: 1,681 assigned augmented protein and 1,716 usual protein. Primary outcomes were available for all retained participants. Median age was 61 years, approximately 64% were male, and median APACHE II score was nineteen. Roughly four-fifths were invasively ventilated, just over half received vasoactive treatment, and 241 had started new kidney replacement therapy before trial feeding. Approximately two-thirds had medical admissions.

Trial feeding began a median of about nineteen hours after intensive care admission. Median exposure was 87 hours with augmented protein and 84 hours with usual protein. Daily formula volumes and calories were similar, while protein delivery separated. Excluding the enrolment day, median recorded delivery was 1.33 versus 0.84 g/kg ideal body weight/day, or 1.04 versus 0.64 g/kg actual body weight/day. These weight-adjusted summaries had missing anthropometric data and should not be treated as complete observations in every participant. Parenteral nutrition was used in approximately 7% of each group and separate protein supplements in less than 1%. Audits supported good allocation adherence.⁵

Median days free of the index hospital and alive at day 90 were 62 versus 64. The adjusted difference was -1.97 days, with a 95% confidence interval of -7.24 to 3.30 and $P=0.46$. The linear analysis gave a mean difference of -1.26 days (95% CI -3.59 to 1.06). The Bayesian analysis estimated a median difference of -1.50 days (95% credible interval -3.86 to 0.90), with a 10.9% posterior probability of any benefit. These were alternative analyses of the same participants.⁴

At ninety days, 1,221/1,681 augmented-protein patients and 1,269/1,716 usual-protein patients were alive, 72.6% versus 74.0%; the adjusted survival risk ratio was 0.99 (95% CI 0.95–1.03). Hospital-free days among survivors were similar. The adjusted mean ventilation-duration difference was 6.8 hours (95% CI -3.0 to 16.5). New kidney replacement therapy occurred in 7.3% versus 7.4%. There was no consistent secondary-outcome advantage. Day-ten median blood urea was higher with augmented protein, 13.0 versus 10.6 mmol/L, among those still being measured in intensive care.

Interaction analyses suggested different effects according to ventilation and preceding kidney replacement therapy. The renal subgroup was added during the trial following publication of EFFORT Protein; it identified dialysis-treated patients rather than all acute kidney injury. Within that subgroup the estimated median difference was -13.19 days, with a wide confidence interval of -50.0 to 23.62 . The investigators rated its effect-modification credibility as moderate. Four adverse events and two serious adverse events were specifically reported; routine outcome events were not systematically counted again as adverse events.^{4,5}

Critique

TARGET Protein gives no affirmative basis for routinely replacing the usual feed with a more protein-dense formula early in critical illness. The principal estimate favoured usual protein, and there was no coherent improvement in survival, ventilation or recovery outcomes. This is a clinically useful conclusion about a feasible feeding policy. It is more limited than showing that protein has no physiological role, that every higher dose is harmful, or that the two strategies are equivalent.

The achieved treatment contrast is central. Formula concentrations differed substantially, delivery audits supported adherence, and recorded protein intake separated while calories remained similar. The intervention therefore cannot be dismissed simply because some patients received less than a guideline target. A randomised comparison can answer whether increasing delivery helps even when neither group consistently reaches an aspirational prescription. Equally, TARGET Protein does

not establish the effect of maintaining every patient at a specified dose for several weeks.

The correspondence usefully exposed how different denominators change the apparent dose. Berris and Elango argued that protein and energy delivery were inadequate and that the amino-acid composition might matter. The investigators replied that actual body weight can overstate requirements in an overweight population and presented delivery relative to ideal weight. Both positions highlight genuine uncertainty about individual targets. However, the letter's overlapping interquartile ranges do not show absence of a treatment difference: they describe between-patient variability, not confidence intervals for the between-group effect. The recorded dose separation remains important evidence.^{6,7}

Timing further qualifies the comparison. Including partial first feeding days lowers average daily delivery, so the 1.33 versus 0.84 g/kg ideal-weight figures must retain their exclusion of the enrolment day. The corresponding summaries across all feeding days were 1.12 versus 0.69 g/kg. Anthropometric measurements were incomplete, and nutrition was recorded daily initially but only intermittently later. These limitations affect precision of dose description; they do not invalidate allocation to the formula policy. Presenting one unqualified grams-per-kilogram figure as the dose received by the whole cohort would conceal these distinctions.⁵

The cluster design facilitated broad participation and an intervention close to ordinary procurement practice. Each unit crossed between both formulae, reducing confounding by stable institutional characteristics. Nevertheless, the randomised units were eight intensive care units, not 3,397 independent treatment allocations. Period adjustment and a unit-level random effect were necessary because case mix, discharge practices and feeding policies could vary within and between units. A large patient count cannot eliminate uncertainty arising from a small number of clusters.

Advance disclosure of the sequence and open treatment also permit selection and performance effects. Clinicians decided whether a formula was contraindicated and how to deliver feeding while knowing the unit's assignment. Broad eligibility, high inclusion and objective ascertainment of death reduce some concerns, but cannot prove that treatment decisions were unaffected. The crossover design controls stable differences more convincingly than evolving period-specific differences. The appraisal should recognise those protections and limitations without treating either the cluster label or the pragmatic label as a verdict on validity.

The acknowledged planning error weakens a simple claim that the trial was powered to exclude a one-day benefit. More decisively, the primary interval still includes an improvement of roughly three days and deterioration exceeding seven days. The appropriate inference comes from that interval and the consistency of related results, not retrospective power alone. The narrower linear and Bayesian estimates are informative sensitivity analyses, but measure or model the outcome differently; they do not replace the designated primary comparison or constitute independent confirmations.

The outcome combines survival and hospital use in a way that is understandable to patients. Assigning zero to death prevents an early death from appearing favourable simply because it shortens admission. However, zero also applies to a survivor remaining in hospital for the whole period, and the distribution had a substantial zero spike. A median contrast does not describe a two-day loss experienced by a typical treated individual. Survival, discharge and readmission need to be considered alongside the composite, as the investigators did.

The phrase hospital-free also needs its qualification. A transfer to another acute hospital counted differently from remaining in the index hospital, even though both involve continuing inpatient care. Approximately one-fifth were discharged to another acute hospital, so this is not merely theoretical. If transfer patterns differed by allocation, the primary outcome could change without an equivalent change in recovery. Local discharge pathways are partly addressed by the cluster crossover design, but the endpoint remains days outside one hospital rather than days at home or free of all hospital care.

Exposure lasted only about three and a half days at the median. This is appropriate to the broad population enrolled, but it dilutes an effect confined to prolonged feeding. Conversely, selecting only people who subsequently remained in intensive care would sacrifice randomisation, because prolonged stay is itself related to outcome. The trial leaves a legitimate question about later rehabilitation and phase-specific protein needs; it does not supply a favourable subgroup merely by excluding rapid recoverers after allocation. Muscle strength, muscle mass and long-term health-related quality of life were not measured.

The renal signal warrants caution rather than a precise treatment rule. Its addition was prompted by external evidence during trial conduct, not presented as an original subgroup. The available definition captured new dialysis before feeding and missed less severe acute kidney injury. Although the reported interaction was strong, the effect estimate within dialysis recipients was highly imprecise. It should not be recast as proof

that each additional protein dose harms every patient with kidney injury. In a trial without overall benefit, it does make confident routine escalation in this group harder to justify.

Higher urea supports a metabolic consequence of greater protein delivery, but does not establish its mechanism of clinical harm. Urea is also influenced by renal clearance and catabolism. Day-ten observations were restricted to patients still available for measurement, a selected subset of the original cohort. Similarly, proposed effects on autophagy or amino-acid balance were not directly tested. These are plausible questions for subsequent work, not explanations demonstrated by the trial.

Safety reporting was deliberately selective. The small number of designated adverse events cannot establish that the formulas were equally safe, particularly when deaths and organ-support outcomes were ordinarily captured elsewhere. The outcome data provide the more useful clinical safety context. The practical judgement is to avoid assuming that more enteral protein early in a general intensive care population improves recovery, while preserving uncertainty about carefully defined later phases, nutritional phenotypes and rehabilitation strategies.

Body of Evidence

The earlier TARGET energy trial illustrates why individual nutritional components require separate evaluation. It analysed 3,957 ventilated adults allocated to energy-dense or routine formula, producing approximately 601 kcal/day more energy with the denser feed. Ninety-day mortality was 26.8% versus 25.7%, relative risk 1.05 (95% CI 0.94–1.16), without an improvement in other major recovery outcomes. That trial did not settle the protein question. It did challenge the assumption that delivering more of a prescribed macronutrient necessarily converts into better clinical recovery.⁸

The protein feasibility trial then tested the same 100 versus 63 g/L formula approach used in TARGET Protein. It randomised 120 ventilated patients in six units, with data available for 116. Mean protein delivery was 1.52 versus 0.99 g/kg ideal body weight/day, a difference of 0.53 (95% CI 0.38–0.69), without an important energy difference. Feeding lasted about eight to nine days. The study established achievable dose separation and masking; its small sample and wide mortality interval could not determine patient benefit. The shorter exposure in the definitive pragmatic trial reflects its broader eligibility and different delivery context.¹

EFFORT Protein supplied a larger, internationally distributed patient-level comparison. Among 1,329 randomised patients, 1,301 were analysed across 85 units in sixteen

countries. Nutritionally high-risk ventilated adults were prescribed at least 2.2 versus no more than 1.2 g/kg/day. Live hospital discharge by sixty days was not improved: hazard ratio 0.91 (95% CI 0.77–1.07). Mortality was 34.6% versus 32.1%, relative risk 1.08 (95% CI 0.92–1.26). Its acute-kidney-injury signal prompted particular concern, but arose within a trial whose overall mortality comparison was uncertain. EFFORT tested a higher prescribed dose in a more selected population than TARGET Protein.²

PRECISE moved the question towards longer-term recovery. In 935 patients across ten Dutch and Belgian hospitals, blinded allocation to higher or standard isocaloric enteral protein did not improve function. The repeated health-utility outcome over 180 days was worse with higher protein, mean difference –0.05 (95% CI –0.10 to –0.01). Gastrointestinal intolerance was more frequent, while the mortality hazard ratio of 1.14 had a confidence interval of 0.92–1.40. This offers a patient-centred outcome absent from TARGET Protein and makes a hidden broad functional benefit less persuasive. The two trials still differ in selection, intended dose and outcome ascertainment.³

A 2024 systematic review of twenty-three randomised trials, comprising 3,303 participants, found no overall mortality advantage from higher protein, relative risk 0.99 (95% CI 0.88–1.11). It identified higher urea concentrations and a concerning mortality result in acute-kidney-injury subgroups. Its search ended in May 2023, so it preceded PRECISE and TARGET Protein. Some contributing patients were already represented in EFFORT; the pooled findings should not be counted as a separate replication of that trial. The mixture of doses and populations also limits any claim to have identified one universally optimal prescription.⁹

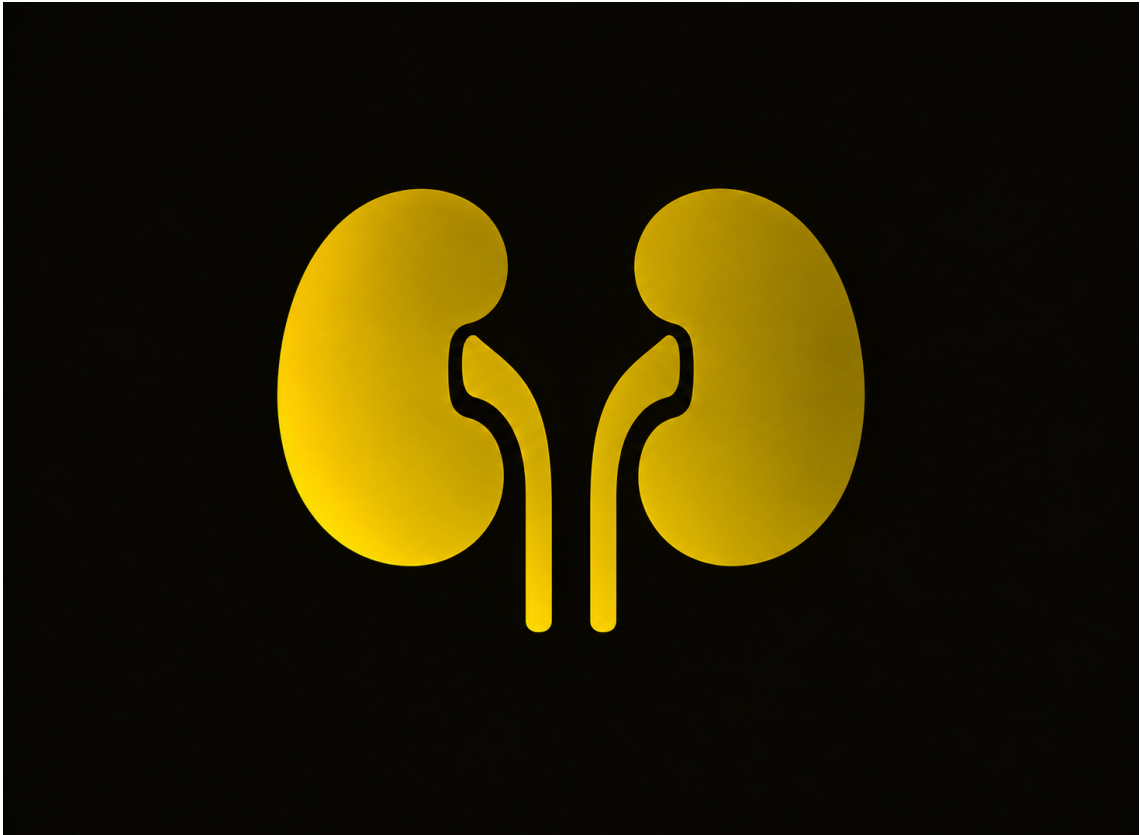
Together, these studies shift the burden of proof. Earlier observational associations and biochemical reasoning supported a hypothesis worth testing, but contemporary randomised comparisons do not support routine early escalation simply to reach a higher protein target. TARGET Protein extends that conclusion to a broad, readily implemented formula policy. It leaves a more focused research agenda around timing, recovery, rehabilitation and patient selection, rather than evidence for indiscriminate protein restriction or abandonment of enteral nutrition.

Should we routinely use a higher-protein enteral formula early in critical illness?

Not routinely. TARGET Protein found no improvement in recovery and adds to concerns about assuming benefit from early protein escalation, particularly in patients receiving kidney replacement therapy.

References

1. Chapple LS, Summers MJ, Bellomo R, Chapman MJ, Davies AR, Ferrie S, et al. Use of a High-Protein Enteral Nutrition Formula to Increase Protein Delivery to Critically Ill Patients: A Randomized, Blinded, Parallel-Group, Feasibility Trial. *JPEN J Parenter Enteral Nutr.* 2021;45(4):699-709.
2. Heyland DK, Patel J, Compher C, Rice TW, Bear DE, Lee ZY, et al. The effect of higher protein dosing in critically ill patients with high nutritional risk (EFFORT Protein): an international, multicentre, pragmatic, registry-based randomised trial. *Lancet.* 2023;401(10376):568-576.
3. Bels JLM, Thiessen S, van Gassel RJJ, Beishuizen A, De Bie Dekker A, Fraipont V, et al. Effect of high versus standard protein provision on functional recovery in people with critical illness (PRECISE): an investigator-initiated, double-blinded, multicentre, parallel-group, randomised controlled trial in Belgium and the Netherlands. *Lancet.* 2024;404(10453):659-669.
4. Summers MJ, Chapple LS, Karahalios A, Bellomo R, Chapman MJ, Ferrie S, et al. Augmented Enteral Protein During Critical Illness: The TARGET Protein Randomized Clinical Trial. *JAMA.* 2025;334(4):319-328.
5. Summers MJ, Chapple LS, Karahalios A, Bellomo R, Chapman MJ, Ferrie S, et al. Supplement 3: supplementary methods and results to Augmented Enteral Protein During Critical Illness: The TARGET Protein Randomized Clinical Trial. *JAMA.* 2025.
6. Berris K, Elango R. Augmented Enteral Protein in Critical Illness. *JAMA.* 2025;334(18):1675.
7. Summers MJ, Chapple LS, Deane AM. Augmented Enteral Protein in Critical Illness-Reply. *JAMA.* 2025;334(18):1675-1676.
8. Chapman M, Peake SL, Bellomo R, Davies A, Deane A, Horowitz M, et al. Energy-Dense versus Routine Enteral Nutrition in the Critically Ill. *N Engl J Med.* 2018;379(19):1823-1834.
9. Lee ZY, Dresen E, Lew CCH, Bels J, Hill A, Hasan MS, et al. The effects of higher versus lower protein delivery in critically ill patients: an updated systematic review and meta-analysis of randomized controlled trials with trial sequential analysis. *Crit Care.* 2024;28(1):15.



Renal Trials

BICARICU-2

Jung B, Jabaudon M, De Jong A, Bitker L, Audard J, Klouche K, et al. Sodium Bicarbonate for Severe Metabolic Acidemia and Acute Kidney Injury: The BICARICU-2 Randomized Clinical Trial. JAMA. 2025;334(22):2000–2010.

Introduction

Severe metabolic acidaemia accompanies many forms of critical illness. It may reflect lactate accumulation during shock, impaired renal acid excretion, chloride administration or several processes at once. A low arterial pH identifies serious physiological disturbance, but the implications of correcting that value depend on the cause. Raising pH does not necessarily reverse impaired perfusion, cellular injury or the disease generating the acid load.

Sodium bicarbonate is readily available and can increase extracellular pH. Its administration also adds sodium and fluid, generates carbon dioxide and can lower ionised calcium. These effects make the balance context-dependent, particularly in patients with limited ventilation or fluid tolerance. The intervention is therefore more than replacement of a laboratory deficit: it changes several physiological variables simultaneously.^{1,2}

The original BICAR-ICU trial did not improve its primary outcome in the overall population with severe metabolic acidaemia. A prespecified subgroup with moderate or severe acute kidney injury appeared to have better survival and less use of kidney replacement therapy. That finding was influential because impaired renal acid clearance supplies a plausible reason for bicarbonate to be more useful in this subgroup. It nevertheless required confirmation in a trial designed specifically for that population.³

BICARICU-2 provided that test. It asked whether a bicarbonate strategy targeting pH at least 7.30 improved 90-day survival in patients with severe metabolic acidaemia and stage 2 or 3 acute kidney injury. It also examined a distinct, clinically important possibility: bicarbonate might avert or postpone kidney replacement therapy without materially changing survival.

Synopsis

BICARICU-2 was an investigator-initiated, open-label randomised trial conducted in 43 French ICUs between October 2019 and December 2023. There was no industry involvement. An independent monitoring board reviewed safety and planned interim analyses. The target of 640 randomised patients was reached, and the principal report was published on 29 October 2025.^{1,4}

Adults were enrolled within 48 hours of ICU admission. Eligibility required pH 7.20 or lower, bicarbonate concentration 20 mmol/L or lower, PaCO₂ 45 mmHg or lower, and stage 2 or 3 acute kidney injury according to KDIGO criteria. The relevant measurements had to be obtained within six hours before enrolment. Patients also needed a SOFA score above four or lactate at least 2 mmol/L, identifying a population with significant acute illness.

Important exclusions were bicarbonate loss from the gastrointestinal or urinary tract, ketoacidosis, selected exogenous acid poisonings, advanced pre-existing kidney failure, recent bicarbonate treatment and kidney replacement therapy already given or planned within six hours. These restrictions matter: the trial addressed severe acidaemia accompanying critical illness and acute kidney injury, rather than every indication for bicarbonate or urgent dialysis.

Central computer allocation used minimisation by centre, age and baseline pH. Patients were assigned equally to intravenous 4.2% sodium bicarbonate or no bicarbonate infusion. There was no placebo. The intervention targeted arterial pH at least 7.30 until day 28 or ICU discharge. Recommended individual infusions were 125–250 mL over thirty minutes, with no more than 1,000 mL during the first 24 hours and repeat blood-gas assessment after administration.

Recommendations for kidney replacement therapy were common to both groups. Immediate treatment was advised for severe hyperkalaemia with electrocardiographic changes or severe anuric pulmonary oedema meeting the specified oxygenation criteria. At 24 hours, treatment was also recommended when at least two of three features persisted: marked oliguria, pH no higher than 7.20 despite resuscitation and potassium above 6.5 mmol/L. Decisions remained clinical, making these recommendations less rigid than automatic treatment triggers.

The primary outcome was all-cause mortality at day 90. Secondary outcomes included mortality at days 28 and 180, use and duration of organ support, ICU and hospital stay, fluid balance, infection and major adverse kidney events. The sample size assumed

approximately 80% control mortality and sought a ten-percentage-point absolute reduction. Secondary confidence intervals were not adjusted for multiple comparisons.

Of 640 randomised patients, 627 were included in the primary analysis, with 314 assigned bicarbonate and 313 control. Thirteen were excluded after allocation for consent, legal eligibility or duplicate enrolment reasons. Day-90 mortality was available for all retained patients. Median age was 67 years, roughly three-quarters were invasively ventilated and approximately four-fifths received vasopressors. Septic shock was the principal attributed cause of acidaemia in just over half.

Baseline pH was 7.15 in both groups. Median bicarbonate concentration was 12–13 mmol/L and lactate approximately 5.7–5.9 mmol/L. The bicarbonate group received a median of 750 mL of 4.2% solution over the first 48 hours, beginning a median of fifteen minutes after enrolment. Arterial pH and bicarbonate increased more with treatment. Forty-seven control patients subsequently received bicarbonate, while two allocated to treatment did not receive it.

Ninety-day mortality was 195/314 (62.1%) with bicarbonate and 193/313 (61.7%) with control, an absolute difference of 0.4 percentage points (95% CI –7.2 to 8.0; $P=0.91$). The mortality hazard ratio was 0.97 (95% CI 0.80–1.19). Adjusted and sensitivity analyses did not establish benefit. Mortality at days 28 and 180 was also not clearly different.

Kidney replacement therapy was used by day 28 in 109/314 patients (34.7%) versus 157/313 (50.2%), an absolute difference of –15.5 percentage points (95% CI –23.1 to –7.8). Among those receiving it, median time to initiation was approximately 31 versus sixteen hours. Persistent acidaemia was a less frequent recorded indication in bicarbonate recipients. Median kidney-replacement-free days were zero in both full groups, reflecting the high mortality burden.

No clear improvement was established in duration of ventilation, vasopressor-free days or length of stay. Major adverse kidney events at day 90 occurred in 199/240 patients (83%) versus 195/232 (84%) with available data. Bloodstream infection was less frequent, but this was one of numerous secondary comparisons. Fluid balance at day two was more positive with bicarbonate, and the trial did not identify a clear overall difference in recorded adverse events.^{1,4}

Critique

BICARICU-2 did not confirm a survival advantage from bicarbonate in the acute kidney injury population highlighted by BICAR-ICU. It is a useful example of directly testing a plausible subgroup hypothesis. The result should reduce confidence in a large mortality benefit, while preserving uncertainty about smaller effects and recognising the separate reduction in kidney replacement therapy.

The mortality confidence interval is more informative than the very high P value. It remains compatible with approximately seven fewer or eight more deaths per hundred patients. The study therefore does not establish mortality equivalence or prove that bicarbonate has no effect. It does, however, make the large benefit anticipated from the preceding subgroup much less persuasive. A neutral point estimate in a dedicated trial should carry more weight than enthusiasm generated by a smaller stratum of an earlier study.

The power assumptions were optimistic. Observed control mortality was approximately 62%, substantially below the assumed 80%, and a ten-percentage-point absolute survival benefit is large for a buffering intervention within complex critical illness. Smaller benefits would require a much larger trial. This limitation explains residual uncertainty; it should not be used to reinterpret a result centred near no effect as evidence that a benefit probably exists. The accompanying editorial appropriately emphasises both the imprecision and the need for more realistic future effect sizes.⁵

The kidney replacement result is compelling in magnitude but requires careful interpretation. Bicarbonate directly treats acidaemia, which was itself a recommended trigger for replacement therapy. Fewer patients crossing that trigger is a plausible intended effect. It does not necessarily mean less structural kidney injury or faster renal recovery. The intervention may substitute temporary buffering for extracorporeal acid correction while the underlying illness resolves.

That distinction does not make the outcome unimportant. Avoiding vascular access, anticoagulation, treatment-related instability and resource use can matter to patients and services. Nevertheless, the benefit is avoidance of a treatment decision under this trial's conditions, not proven nephroprotection. The neutral major adverse kidney event comparison and absence of a clear improvement in other organ support outcomes prevent a stronger claim about recovery.

Open allocation also permits clinician expectations to affect when kidney replacement begins. Common recommendations reduce variation, but they do not eliminate judgement about whether acidaemia is refractory or another period of observation is

acceptable. Clinicians aware that bicarbonate is being tested to avert replacement therapy may tolerate a longer delay. The observed reduction can therefore combine genuine physiological substitution with altered decision thresholds; this design cannot precisely separate them.

The timing analysis is conditional on receiving replacement therapy. The patients treated in each group are no longer comparable simply because their original allocation was random. Bicarbonate may remove some lower-risk patients from the treated subset, leaving a different selection of patients who still require therapy. Median initiation at 31 rather than sixteen hours is descriptive of those selected recipients, not an unbiased estimate that every patient can safely wait another fifteen hours.

Similarly, comparing mortality only among patients who received kidney replacement therapy conditions on an outcome affected by treatment and illness severity. Such a comparison cannot establish that delayed initiation is harmless. The relevant randomised evidence is the full-group mortality comparison, whose confidence interval remains broad. Urgent indications for replacement therapy were specifically recognised, and patients already requiring imminent treatment were excluded.

Treatment crossover is another limitation. Fifteen per cent of controls received bicarbonate, potentially reducing separation between policies. Those patients had high mortality and frequent replacement therapy, consistent with rescue use during deterioration. Reassigning them according to treatment received therefore introduces confounding by indication. The analysis described as per-protocol is largely an as-treated comparison, with further exclusions, rather than a clean randomised estimate of adherence.

Post hoc inverse-probability weighting attempts to address measured predictors of treatment deviation, but cannot remove unmeasured or inaccurately recorded reasons for rescue treatment. Its more favourable point estimate should not displace the prespecified allocation-based result. The practical trial question includes the possibility that clinicians will administer bicarbonate despite assignment to a no-routine-bicarbonate policy when circumstances change.

Multiplicity limits the interpretation of other secondary findings. The lower bloodstream infection proportion might relate to less extracorporeal treatment or vascular access, but mediation was not established. Many outcomes and subgroups were examined without adjusted intervals. The same table also contains nominally unfavourable differences, including greater vasopressor exposure. Selecting only favourable intervals

would create an unbalanced account. None supplies a secure additional benefit or harm claim on its own.

Fluid administration is part of the intervention. A median 750 mL of 4.2% bicarbonate supplies a substantial sodium and volume load, and fluid balance was more positive by day two. That difference should not be treated as unrelated confounding arising outside randomisation: it is partly a consequence of the assigned strategy. It may nevertheless influence urine output, haemodynamics and the threshold for replacement therapy, complicating a purely acid–base explanation of the result.

The absence of a detected adverse-event difference should be interpreted within the enrolled population and monitoring arrangements. Most participants were ventilated, and patients with important alternative causes of acidosis were excluded. The trial does not establish safety for unrestricted doses, inadequate carbon dioxide clearance or severe fluid intolerance. It tested repeated, monitored administration within a defined clinical strategy.

Heterogeneity of acidaemia remains important. Similar pH values can arise from different mixtures of lactate, chloride and retained unmeasured anions. BICARICU-2 was dominated by shock and elevated lactate, rather than isolated failure of renal acid excretion. Exploratory categorisation of acidosis and subgroup analyses did not establish a reliable selection rule. A suggestive estimate in the lowest-pH stratum is insufficient to define a proven benefiting subgroup without a convincing interaction and replication.

Follow-up quality also varies by endpoint. Mortality was complete in the primary population, a major strength. The major adverse kidney event outcome had data for 472 of 627 participants, limiting its ability to exclude a renal recovery difference. Missing measurements were not imputed. Conclusions about long-term kidney outcomes should consequently be more cautious than the statement that the primary mortality analysis was complete.

There are reporting details requiring editorial caution. The flowchart caption describes certain blood-gas and kidney-stage outliers as excluded from the primary analysis, while the displayed counts place these exclusions in the additional treatment-received analysis. The chapter follows the consistent primary denominators of 314 and 313 and flags the discrepancy separately. Such inconsistencies do not change the quoted mortality result, but should not be silently reproduced as compatible descriptions.

The practical inference is bounded. Bicarbonate may provide a monitored means of managing severe metabolic acidaemia in selected patients with acute kidney injury and

no immediate replacement indication. Its use should not be justified as a proven survival treatment, nor should a transient improvement in pH obscure continuing shock, hyperkalaemia, pulmonary oedema or another urgent problem. The trial supports distinguishing physiological management from claims of improved overall prognosis.

Body of Evidence

Small physiological studies had already challenged the assumption that correcting arterial pH necessarily improves circulation. In a randomised crossover study of fourteen patients with lactic acidosis, Cooper and colleagues compared bicarbonate with equimolar sodium chloride. Bicarbonate increased pH, raised PaCO₂ and lowered ionised calcium, but did not improve haemodynamic responses relative to saline. The study was far too small and short to address survival, yet demonstrates why a favourable blood-gas change cannot serve as a clinical outcome.²

BICAR-ICU moved to patient-centred outcomes in 389 critically ill adults with severe metabolic acidaemia. Its composite of death by day 28 or organ failure at day seven occurred in 66% with bicarbonate and 71% with control, an absolute difference of –5.5 percentage points (95% CI –15.2 to 4.2). The overall result was not significant. In the prespecified acute kidney injury stratum, estimated day-28 survival was 54% versus 37%, generating the hypothesis subsequently tested by BICARICU-2. Hypernatraemia, hypocalcaemia and metabolic alkalosis were more frequent with bicarbonate.³

The larger dedicated follow-up trial therefore changes the interpretation of that subgroup. Differences in populations, care and endpoint timing may contribute to different estimates, but chance exaggeration in a small subgroup remains a straightforward explanation. BICARICU-2 did reproduce a reduction in replacement therapy, suggesting greater consistency for a treatment-substitution effect than for survival.

Kidney replacement timing trials provide related evidence about the value of avoiding unnecessary initiation. AKIKI randomised 620 severely ill patients with stage 3 acute kidney injury who lacked an immediately life-threatening renal complication. Mortality did not differ clearly between early and delayed strategies, while 49% in the delayed group never received replacement therapy. Catheter-related bloodstream infection was less frequent with delay. This supports careful observation in selected patients, but does not itself test bicarbonate.⁶

STARRT-AKI provided a larger international comparison. Among 2,927 analysed patients, accelerated replacement therapy increased exposure from 61.8% to 96.8% without

reducing 90-day mortality, which was approximately 44% in both groups. Adverse events were more frequent, and surviving accelerated-strategy patients were more often still dependent on replacement therapy at day 90. These findings strengthen the rationale for avoiding treatment without a clear indication, while preserving the need to intervene when urgent complications arise.⁷

A 2025 target trial emulation examined bicarbonate in 6,157 eligible admissions across twelve Australian ICUs. It associated treatment with a 1.9-percentage-point reduction in 30-day ICU mortality. That smaller effect is compatible with the uncertainty in BICARICU-2, but the observational design, broader pH eligibility and different outcome prevent direct equivalence. Modelling and careful treatment-timing methods reduce some biases without reproducing randomisation or guaranteeing control of confounding.⁸

Taken together, the evidence makes a large survival benefit from routine bicarbonate less likely than the original subgroup suggested. It leaves a credible role for temporary acid–base management that may avert kidney replacement therapy in some patients. Whether this produces smaller survival gains, better renal recovery or worthwhile resource savings requires more precise trials with clear treatment thresholds and complete follow-up. Those questions should remain separate from the established ability of bicarbonate to raise arterial pH.

Does bicarbonate improve survival in severe metabolic acidaemia with moderate or severe acute kidney injury?

BICARICU-2 did not demonstrate a survival benefit. It reduced kidney replacement therapy use, supporting selective monitored use to manage acidaemia while recognising that improved pH is not proof of renal recovery or improved prognosis.

References

1. Jung B, Jabaudon M, De Jong A, Bitker L, Audard J, Klouche K, et al. Sodium Bicarbonate for Severe Metabolic Acidemia and Acute Kidney Injury: The BICARICU-2 Randomized Clinical Trial. *JAMA*. 2025;334(22):2000-2010.
2. Cooper DJ, Walley KR, Wiggs BR, Russell JA. Bicarbonate does not improve hemodynamics in critically ill patients who have lactic acidosis. A prospective, controlled clinical study. *Ann Intern Med*. 1990;112(7):492-498.

3. Jaber S, Paugam C, Futier E, Lefrant JY, Lasocki S, Lescot T, et al. Sodium bicarbonate therapy for patients with severe metabolic acidaemia in the intensive care unit (BICAR-ICU): a multicentre, open-label, randomised controlled, phase 3 trial. *Lancet*. 2018;392(10141):31-40.
4. Jung B, Jabaudon M, De Jong A, Bitker L, Audard J, Klouche K, et al. BICARICU-2 supplemental online content. *JAMA*. 2025. doi:10.1001/jama.2025.20231.
5. Kusirisin P, Zampieri FG, Bagshaw SM. Sodium Bicarbonate in Severe Acidemia and Acute Kidney Injury—Turning the Tide or Chasing a Myth? *JAMA*. Published online October 29, 2025.
6. Gaudry S, Hajage D, Schortgen F, Martin-Lefevre L, Pons B, Boulet E, et al. Initiation Strategies for Renal-Replacement Therapy in the Intensive Care Unit. *N Engl J Med*. 2016;375(2):122-133.
7. Bagshaw SM, Wald R, Adhikari NKJ, Bellomo R, da Costa BR, Dreyfuss D, et al. Timing of Initiation of Renal-Replacement Therapy in Acute Kidney Injury. *N Engl J Med*. 2020;383(3):240-251.
8. Blank SP, Blank RM, Laupland KB, Tabah A, Gill D, Kumar A, et al. Sodium bicarbonate administration for metabolic acidosis in the intensive care unit: a target trial emulation. *Intensive Care Med*. 2025;51(6):1078-1086.

LIBERATE-D

Liu KD, Siew ED, Tuot DS, Vijayan A, Matzumura Umemoto G, Birkelo BC, et al. A Conservative Dialysis Strategy and Kidney Function Recovery in Dialysis-Requiring Acute Kidney Injury: The Liberation From Acute Dialysis (LIBERATE-D) Randomized Clinical Trial. JAMA. 2026;335(4):326-335. Published online 7 November 2025.

Introduction

Starting dialysis for acute kidney injury establishes a treatment schedule that may outlast the indication for each individual session. After haemodynamic and respiratory stabilisation, a patient can remain on thrice-weekly haemodialysis while clinicians await convincing evidence of recovery. Yet the treatment itself changes serum creatinine, urea, fluid balance and urine output, making native kidney function difficult to assess. Continuing a familiar schedule may therefore delay the opportunity to discover that dialysis is no longer needed.

There is also a plausible biological concern. Intermittent haemodialysis exposes recovering kidneys to circulatory stress, and removing fluid can cause hypotension even when blood pressure was satisfactory before treatment. Whether reducing unnecessary sessions improves kidney repair, merely reveals recovery earlier, or does both is clinically important. Avoiding treatment carries competing risks from fluid accumulation, uraemia, hyperkalaemia and acidosis.

Trials of dialysis intensity and initiation have challenged the assumption that greater or earlier extracorporeal support improves outcomes. They have not directly answered how to withdraw treatment from a stable patient who has already started it. LIBERATE-D addressed this later decision, comparing indication-driven dialysis with a conventional thrice-weekly schedule in patients recovering from dialysis-requiring acute kidney injury.¹

Synopsis

LIBERATE-D was an open-label, individually randomised superiority trial at four United States clinical sites, including two hospitals within the San Francisco site. Recruitment took place from January 2020 to March 2025, with follow-up completed in June 2025. The National Institutes of Health funded the work; investigators stated that the funders had no role in conduct, analysis or publication. Although its journal issue appeared in January 2026, the principal report was published online in November 2025 and falls within this book's evidence period.

Eligible adults had acute kidney injury attributed at least partly to acute tubular necrosis, had received at least one calendar day of continuous kidney replacement therapy or an intermittent haemodialysis treatment, and had further haemodialysis planned. They had to be haemodynamically stable without vasopressors. The main methods specified a pre-illness estimated glomerular filtration rate of at least 15 mL/min/1.73 m². Important exclusions included endotracheal ventilation, substantial oxygen requirements, prolonged dialysis before enrolment, advanced heart failure requiring continuous inotropes or a ventricular assist device, and anticipated discharge within forty-eight hours. Clinical equipoise and consent were required.¹

Baseline kidney function was calculated from available outpatient creatinine measurements. An amendment described in the main report permitted enrolment without these measurements in patients younger than 65 years without known diabetes, hypertension or chronic kidney disease. Eight participants met this route. The study retained its original 2009 creatinine-based estimating equation for consistency, rather than switching equations during recruitment.

A central system assigned patients equally using permuted blocks, stratified by clinical site and baseline estimated filtration above or below 45 mL/min/1.73 m². Enrolling personnel did not know the sequence or block size. Treatment and outcome assessment were unblinded. The conservative group received dialysis for specified metabolic or clinical indications, or when the treating clinician judged that it was necessary. The strategy was therefore a supervised decision framework, not a prohibition on treatment until rigid thresholds were crossed.

Specified triggers included serum urea nitrogen above 112 mg/dL, equivalent to approximately 40 mmol/L urea; potassium above 6 mmol/L, or above 5.5 mmol/L despite medical treatment; and severe metabolic acidosis. Pulmonary oedema with substantial oxygen requirements despite diuretics also qualified, with a lower increase-in-oxygen threshold for anuric patients. Pure ultrafiltration was recommended when volume overload was the sole indication. Clinicians could return deteriorating patients to continuous therapy.^{1,2}

The control strategy scheduled haemodialysis three times weekly until recovery criteria were reached. These included timed creatinine clearance above 20 mL/min, with clinician discretion between 8 and 20 mL/min; urine output above 1 L/day without diuretics or 2 L/day with diuretics; or a spontaneous creatinine fall exceeding 0.3 mg/dL over more than twelve hours. Once dialysis was stopped, the conservative indications governed restarting it. Final treatment decisions remained with the clinical team in both groups.

The intervention ended at discharge from the study hospital. Subsequent care was not controlled, although outcomes were followed through ninety days using records and telephone contact. The primary outcome was being alive and independent of dialysis at hospital discharge, with independence sustained for at least fourteen consecutive days, including days after discharge. A patient who stopped dialysis but subsequently died in hospital did not meet the primary outcome. This definition required evidence of sustained treatment independence rather than a single omitted session.

Two key secondary outcomes were dialysis sessions per week and dialysis-free days to day 28. The latter counted consecutive days alive and independent after the final treatment, not the ordinary gaps between intermittent sessions; death by day 28 contributed zero. For the treatment-frequency measure, a day of continuous therapy counted as three-sevenths of an intermittent session. Other outcomes included recovery and mortality at days 28 and 90, hospital stay, and time to recovery.

The planned sample of 220 provided 80% power for a twenty-percentage-point improvement in recovery, assuming a control recovery rate between 30% and 50%. The primary comparison was unadjusted, using a chi-squared test. A prespecified logistic model additionally adjusted for demographic variables and baseline kidney function. The supplementary report provides both complete-case and imputed-baseline-function versions of that model. An independent monitoring board reviewed the trial, including an interim review after approximately half the planned recruitment.^{1,2}

Of 909 patients assessed, 221 were randomised. One immediately identified ineligible participant was excluded and replaced, leaving 220 who received their allocation, 110 per group. A kidney transplant recipient and a participant who withdrew left 109 per group for the primary comparison. Mean age was 56 years, 67% were male, and mean pre-illness estimated filtration was approximately 65 mL/min/1.73 m². Dialysis had begun a median of nine days earlier. About 40% remained in intensive care, and most had previously needed intensive care or vasopressors, but none was receiving vasopressors at enrolment.

The conservative group received a median of 1.8 dialysis sessions per week versus 3.1 with conventional treatment, a difference of -1.4 (95% CI -1.8 to -1.0). Only 72/110 conservative-group patients received any further dialysis, compared with 101/110 controls. There were 255 versus 505 haemodialysis treatments, but more pure ultrafiltration treatments, 67 versus 33, and 105 versus 68 days of continuous therapy. Clinician judgement triggered 116/255 conservative-group haemodialysis sessions, 45.5%, commonly because of volume or uraemic concerns below the formal thresholds.²

Sustained dialysis independence at discharge occurred in 70/109 patients, 64.2%, versus 55/109, 50.5%. The absolute difference was 13.8 percentage points (95% CI 0.8–26.8; $P=0.04$), and the unadjusted odds ratio was 1.76 (95% CI 1.02–3.03). After prespecified adjustment, the odds ratio was 1.56 (95% CI 0.86–2.84; $P=0.15$). That complete-case model included 210 participants; assigning a preserved baseline filtration estimate to the eight patients without prior measurements gave a similar adjusted result.^{1,2}

Median dialysis-free days to day 28 were 21 versus five, a difference of sixteen days (95% CI 5–27). Recovery at day 28 was 65.1% versus 54.1%, and at day 90 it was 66.7% versus 58.3%; both confidence intervals included no difference. Hospital stay after randomisation was similar. Ninety-day mortality was 14.7% versus 18.5%, an absolute difference of –3.8 percentage points (95% CI –13.7 to 6.1). Day-28 mortality was numerically higher with conservative treatment, 12.7% versus 6.4%, also with substantial uncertainty.

There were 69 versus 97 dialysis-associated hypotension events, affecting 35 versus forty patients. Serious adverse events other than death affected eleven versus thirteen patients, with one conservative-group event considered probably related to participation. Urgent dialysis and new hypoxaemia were numerically more frequent with conservative treatment. No severe hyperkalaemia events meeting the predefined threshold were recorded. These small safety counts did not establish equivalence between strategies.^{1,2}

Critique

The trial addresses a neglected, recognisable decision. Most dialysis trials begin when patients are critically ill and treatment is being initiated. LIBERATE-D instead asks whether the prescription should change as physiology improves. Its individual randomisation, clear separation in treatment exposure and nearly complete primary-outcome data strengthen the comparison. The result is directly relevant to the routine reassessment of established acute dialysis.

Its population requires careful definition. These were not patients with uncontrolled shock or substantial endotracheal ventilatory requirements. Although many had survived severe critical illness and some remained in intensive care, they had crossed a stability threshold before enrolment. The practical boundary is physiological rather than simply the name of the ward. Applying the conservative schedule during ongoing vasopressor-dependent shock would extrapolate beyond the trial.

The primary outcome combines an important patient benefit with a decision that the intervention deliberately changes. Independence from dialysis matters even if achieved by recognising adequate native function sooner. However, the trial cannot cleanly separate earlier recognition of recovery from faster biological repair. The conservative strategy created immediate opportunities to test whether treatment remained necessary, whereas the control strategy required specified recovery signals before abandoning a schedule. The recovery curves separated very early, and more than one-third of conservative-group patients received no further dialysis. Some were probably already able to manage without another session.

That observation does not make the outcome meaningless. Avoiding unnecessary procedures, vascular access exposure and restrictions can be worthwhile without demonstrating accelerated tubular regeneration. It does narrow the causal claim. LIBERATE-D supports a treatment policy that reduced dialysis exposure and increased the observed probability of sustained independence at discharge. It does not establish that dialysis itself caused structural kidney injury or that withholding it directly repaired the kidneys.

Unblinded assessment compounds this distinction. Outcome definitions were explicit, but clinicians knew allocation when interpreting fluid status, symptoms and biochemical trends. The different stopping frameworks were integral to the comparison. Consequently, a seemingly objective endpoint such as dialysis dependence still incorporates clinical judgement. The fourteen-day confirmation requirement reduces the risk of counting a brief unsuccessful pause as recovery, but it cannot remove the influence of treatment decisions on the endpoint.

The primary estimate is promising and imprecise. Its confidence interval ranges from a small absolute benefit to a very substantial one. The study was planned around a twenty-point improvement, larger than the observed 13.8 points. It cannot provide a stable estimate of the benefit that another hospital should expect. The additional adjusted model matters because the sample was small and some baseline characteristics, including age and comorbidity, favoured the conservative group.

Adjustment attenuated the odds ratio and widened the uncertainty across no effect. This does not retroactively replace the prespecified unadjusted primary analysis, nor does $P=0.15$ prove that there is no benefit. Both analyses point in the same direction, but the conventional statistical declaration is sensitive to the analysis. The complete-case model also omitted eight patients without historical creatinine measurements, although the imputed analysis was reassuringly similar. A balanced interpretation preserves the positive primary result while avoiding a claim of definitive superiority.

The large difference in dialysis-free days is consistent with the intervention's practical effect, but it should not be mistaken for a separate biological confirmation. Treatment frequency, time to liberation and dialysis-free days are closely connected. The outcome correctly excludes routine nondialysis days between sessions; nevertheless, a strategy that tests independence earlier can improve it by changing the timing of recognition. Mortality is incorporated through a zero score, making the distribution harder to interpret as a simple duration benefit for an individual patient.

The time-to-recovery summaries deserve similar caution. Reported medians among patients contributing recovery times are not a whole-cohort guarantee of how quickly someone will recover. The displayed curve censored patients at death, which is a competing event rather than an innocuous loss to follow-up. The discharge primary outcome dealt with death by requiring survival, but the curve answers a different question. Neither presentation should be used to imply that all randomised patients recovered more quickly.

The exposure data also resist a simple 'half as much dialysis' account. Haemodialysis treatments approximately halved, while ultrafiltration and continuous-therapy days increased numerically. The composite frequency measure converts modalities using a practical convention rather than a demonstrated equivalence in physiological burden, staff workload or cost. The trial convincingly reduced scheduled intermittent haemodialysis; it did not provide a complete economic evaluation of all resource use.

Clinician discretion was central to implementation. Almost half the conservative haemodialysis treatments occurred without a formal biochemical or respiratory trigger. This is not automatically a protocol failure: discretion was explicitly permitted. It shows that the intervention relied on repeated expert assessment and the ability to respond promptly to deterioration. Removing that judgement and treating the thresholds as mandatory waiting targets would implement a materially different strategy.

The hypotension findings are compatible with less cumulative exposure to haemodialysis, but do not establish safer individual sessions. There were fewer total events and slightly fewer affected patients, alongside many fewer intermittent treatments. These denominators answer different questions. Hypotension during continuous therapy was not captured by the same event definition, and ultrafiltration rates during intermittent treatment were higher in the conservative group. One cannot infer that each conservative-group treatment was intrinsically gentler.

Safety uncertainty remains clinically consequential. Urgent treatments and hypoxaemia were numerically more common, while mortality differences changed direction between

day 28 and day 90. The trial was too small to resolve uncommon harms or moderate mortality effects. The absence of a statistically significant safety difference is not assurance that less frequent dialysis is safe in every recovery trajectory. Close monitoring and prompt rescue were part of the setting in which the findings arose.

Finally, the intervention ended at hospital discharge, and the recovery difference narrowed by day 90. Outpatient prescribing could reinforce or erase the inpatient effect, but was not controlled. A larger study spanning this transition would help distinguish earlier liberation from a durable increase in recovery and could measure symptoms, quality of life, readmissions and total treatment burden. The accompanying editorial argued for a substantial change in routine practice; its strongest mechanistic claims go beyond what this modest trial alone can establish.³

Body of Evidence

The Acute Renal Failure Trial Network study randomised 1,124 critically ill patients to more or less intensive renal support. The less intensive strategy used thrice-weekly intermittent dialysis or lower-intensity continuous treatment, while the more intensive strategy increased both. Sixty-day mortality was 53.6% versus 51.5%, odds ratio 1.09 (95% CI 0.86–1.40), without better kidney recovery. More patients experienced hypotension with intensive treatment. Its conventional intermittent schedule informed subsequent practice, but its much sicker population did not establish that recovering patients require three treatments every week regardless of native clearance.⁴

RENAL studied continuous-treatment intensity in 1,508 critically ill adults, comparing effluent prescriptions of 40 and 25 mL/kg/hour. Among participants with primary-outcome data, ninety-day mortality was 44.7% in both groups. Higher intensity increased hypophosphataemia and did not establish better dialysis independence. Together, these intensity trials weakened the assumption that more clearance necessarily improves clinical outcomes. They did not test LIBERATE-D's indication-driven withdrawal strategy.⁵

AKIKI moved the question to treatment initiation. In 620 patients with severe acute kidney injury requiring ventilation, catecholamines or both, early initiation did not improve sixty-day survival compared with waiting for specified indications. Forty-nine per cent of the delayed group avoided dialysis. This demonstrates that a biochemical stage of kidney injury need not automatically require extracorporeal support when urgent complications are absent. The trial concerned whether to start treatment, rather than how to stop established treatment.⁶

STARRT-AKI provided a larger multinational comparison. Among 2,927 patients in its modified intention-to-treat analysis, accelerated initiation did not reduce ninety-day mortality: 43.9% versus 43.7%, relative risk 1.00 (95% CI 0.93–1.09). Among ninety-day survivors, ongoing dialysis dependence was 10.4% versus 6.0%, relative risk 1.74 (95% CI 1.24–2.43). Adverse events were also more frequent with acceleration. These findings support caution about unnecessary exposure, while the survivor-restricted dependence comparison and different treatment stage prevent direct equivalence with LIBERATE-D.⁷

AKIKI 2 defines an important limit to the 'less treatment' argument. It randomised 278 patients who had already reached prolonged oliguria or substantial uraemia to delayed or still further delayed initiation. Additional postponement did not improve days alive and free of treatment, and the adjusted mortality analysis suggested harm, hazard ratio 1.65 (95% CI 1.09–2.50). Waiting is therefore a strategy with conditions and boundaries, not an inherently protective intervention.⁸

LIBERATE-D adds a distinct recovery-phase question to this evidence. Its findings support actively challenging an automatic dialysis schedule once patients are stable, assessing residual function and current indications at each decision. The evidence is stronger for reducing unnecessary treatments than for a precise improvement in long-term renal recovery. Confirmation in a larger population, with sustained follow-up across discharge and adequately assessed safety, would make the size and durability of the benefit clearer.

Should stable patients recovering from acute kidney injury continue automatic thrice-weekly dialysis?

Reassess the need for each treatment. LIBERATE-D supports a closely monitored, indication-driven strategy in selected stable patients, while leaving uncertainty about the size of the recovery benefit and uncommon harms.

References

1. Liu KD, Siew ED, Tuot DS, Vijayan A, Matzumura Umemoto G, Birkelo BC, et al. A Conservative Dialysis Strategy and Kidney Function Recovery in Dialysis-Requiring Acute Kidney Injury: The Liberation From Acute Dialysis (LIBERATE-D) Randomized Clinical Trial. JAMA. 2026;335(4):326-335. First published online in 2025.

2. Liu KD, Siew ED, Tuot DS, Vijayan A, Matzumura Umemoto G, Birkelo BC, et al. Supplementary methods and results to A Conservative Dialysis Strategy and Kidney Function Recovery in Dialysis-Requiring Acute Kidney Injury: The Liberation From Acute Dialysis (LIBERATE-D) Randomized Clinical Trial. JAMA. Published online 2025.
3. Waikar SS. A-E-I-O-U and Sometimes Why-Dialysis in Acute Kidney Injury. JAMA. 2026;335(4):312-313. First published online in 2025.
4. Palevsky PM, Zhang JH, O'Connor TZ, Chertow GM, Crowley ST, Choudhury D, et al. Intensity of renal support in critically ill patients with acute kidney injury. N Engl J Med. 2008;359(1):7-20.
5. Bellomo R, Cass A, Cole L, Finfer S, Gallagher M, Lo S, et al. Intensity of continuous renal-replacement therapy in critically ill patients. N Engl J Med. 2009;361(17):1627-1638.
6. Gaudry S, Hajage D, Schortgen F, Martin-Lefevre L, Pons B, Boulet E, et al. Initiation Strategies for Renal-Replacement Therapy in the Intensive Care Unit. N Engl J Med. 2016;375(2):122-133.
7. Bagshaw SM, Wald R, Adhikari NKJ, Bellomo R, da Costa BR, Dreyfuss D, et al. Timing of Initiation of Renal-Replacement Therapy in Acute Kidney Injury. N Engl J Med. 2020;383(3):240-251.
8. Gaudry S, Hajage D, Martin-Lefevre L, Lebbah S, Louis G, Moschietto S, et al. Comparison of two delayed strategies for renal replacement therapy initiation for severe acute kidney injury (AKIKI 2): a multicentre, open-label, randomised, controlled trial. Lancet. 2021;397(10281):1293-1300.

BigpAK-2

Zarbock A, Ostermann M, Forni L, Bode C, Wild L, Putensen C, et al. A preventive care strategy to reduce moderate or severe acute kidney injury after major surgery (BigpAK-2); a multinational, randomised clinical trial. Lancet. 2025;406(10521):2782-2791.

Introduction

Postoperative acute kidney injury often emerges after several potentially modifiable insults have accumulated. Hypotension, impaired cardiac output, inappropriate fluid administration and nephrotoxic exposure may coexist with inflammation and underlying vulnerability. A single protective drug is unlikely to address this entire process. Reliable delivery of supportive care could therefore matter as much as discovering a new treatment.

The difficulty is deciding whom to target and when. Creatinine may rise after the opportunity for prevention has narrowed, while treating every surgical patient with repeated cardiac-output measurements and an intensive care bundle would consume substantial resources. Urinary tissue inhibitor of metalloproteinases-2 and insulin-like growth factor-binding protein 7, combined as TIMP-2 × IGFBP7, identify cellular stress associated with subsequent kidney injury. Their predictive performance creates an opportunity to select a higher-risk population before conventional functional criteria are met. Prediction alone, however, does not demonstrate that acting on a test improves outcomes.¹

Small trials of biomarker-selected supportive care had suggested a reduction in postoperative kidney injury. BigpAK-2 extended this approach across surgical specialties and hospitals. It asked whether a structured preventive strategy reduced moderate or severe acute kidney injury within seventy-two hours after major surgery in patients selected by both clinical risk and a positive urinary biomarker.²

Synopsis

BigpAK-2 was an investigator-initiated, open-label, adaptive randomised trial in thirty-four academic and non-academic hospitals across eight European countries. Recruitment ran from November 2020 to June 2024. BioMérieux provided an unrestricted research grant and free biomarker kits. The investigators stated that the funder had no role in study design, data collection, analysis, interpretation or writing. Several investigators disclosed relevant commercial relationships.²

Patients were screened after major surgery expected to last more than two hours and require intensive care or high-dependency admission. Eligible adults had urinary and central venous catheters and at least one specified risk factor: age at least 75 years, ongoing postoperative ventilation or vasopressor support, stage 3 chronic kidney disease, or intraoperative radiocontrast exposure. They also required a urinary TIMP-2 × IGFBP7 value of at least 0.3 (ng/mL)²/1,000. Repeated biomarker measurements were permitted.

The principal exclusion criteria included estimated filtration below 30 mL/min/1.73 m², recent kidney transplantation, anuria, established acute kidney injury and an indication for dialysis. The study consequently addressed prevention in biomarker-positive patients without established functional kidney injury, rather than treatment of advanced acute kidney injury or advanced chronic kidney disease. Consent was obtained from patients or authorised representatives under local arrangements.

The main report describes amendments to the measurement window and surgical mix. Biomarker sampling moved from the first four postoperative hours to a four-to-eighteen-hour window, and recruitment was extended to permit up to 500 cardiac surgical patients. These changes accommodated overnight admissions and broadened the population. Randomisation was web-based, in equal proportions, with computer-generated permuted blocks stratified by site. Patients and treating staff were unmasked, but staff adjudicating the primary outcome were masked to allocation.

Both groups received usual preoperative and intraoperative care. The intervention began postoperatively with at least twelve hours of structured haemodynamic assessment. Passive leg raising was repeated approximately every three hours, with a rise in cardiac output exceeding 10% prompting balanced crystalloid boluses. The algorithm targeted mean arterial pressure of at least 65 mmHg and cardiac index of at least 2.5 L/min/m², using vasoactive or inotropic support when required. Cardiac output could be measured by several accepted techniques, including echocardiography, pulmonary artery catheterisation and less invasive monitoring.^{2,3}

The remaining measures aimed to reduce avoidable renal stress. Angiotensin-converting enzyme inhibitors and angiotensin receptor blockers were withheld for at least forty-eight hours, and other potentially nephrotoxic drugs, radiocontrast, starch, gelatin and chloride-rich fluids were avoided for at least seventy-two hours where possible. A blood glucose target of 100–150 mg/dL, approximately 5.5–8.3 mmol/L, was pursued with insulin when necessary. Clinicians could withhold individual measures judged inappropriate for a particular patient. Control patients received usual postoperative care.

The primary outcome was KDIGO stage 2 or 3 acute kidney injury within seventy-two hours of surgery, defined by serum creatinine or urine output. Preoperative creatinine provided the baseline. Secondary outcomes included any acute kidney injury, persistence of moderate or severe injury, bundle adherence, biomarker change, organ-support-free days, dialysis, mortality, length of stay and major adverse kidney events. The ninety-day composite included death, dialysis and persistent renal dysfunction, the latter defined by creatinine exceeding twice its preoperative value.

The sample-size assumptions were a 20% primary-event rate with usual care and 14% with the intervention. An adaptive design incorporated an interim analysis planned after 618 evaluable participants, with predefined stopping boundaries and sample-size recalculation. Recruitment continued for the additional cohort specified by the interim decision. Stage-specific, centre-stratified tests were combined using an inverse-normal method with equal weights, and repeated confidence intervals accounted for the sequential design. The report states that overall type I error was controlled at 5%. Secondary and subgroup analyses were exploratory without multiplicity adjustment.^{2,3}

Of 7,873 patients screened, 1,180 were randomised, 589 to intervention and 591 to control. Four withdrew consent, leaving 1,176 retained participants; another four lacked primary-outcome data, so the outcome table contains 584 and 588 observations. Median ages were 72 and 71 years, and approximately two-thirds were male. About one-third underwent cardiac surgery, one-third general or abdominal surgery, and 15% vascular surgery. Around 90% of operations were elective. The population therefore had broad surgical representation but was concentrated in older patients undergoing planned major procedures.

Care delivery separated substantially. Advanced haemodynamic monitoring was recorded in 78.6% versus 8.8%, and regular passive leg raising in 80.6% versus 10.5%. Full bundle delivery occurred in 46.9% versus 5.0%; excluding the glucose component increased the intervention proportion to 62.7%. Prolonged hypotension without an intervention to prevent it occurred in 7.6% versus 15.3%. Hyperglycaemia lasting more

than three hours occurred in 29.1% versus 42.3%. Overall fluid and vasopressor exposure differed less markedly, while dobutamine was used more frequently in the intervention group.³

Moderate or severe acute kidney injury occurred in 84/584 intervention patients, 14.4%, and 131/588 controls, 22.3%. The sequentially analysed odds ratio was 0.57 (95% CI 0.40–0.79; $P=0.0002$). The observed absolute difference was approximately eight percentage points. Creatinine-based and urine-output-based analyses both favoured intervention, although urine output accounted for more events: the respective creatinine-based counts were 37 versus 59, and urine-output-based counts 65 versus 100. These categories overlapped.^{2,3}

Any-stage acute kidney injury occurred in 213/584 versus 240/588, 36.5% versus 40.8%, odds ratio 0.78 (95% CI 0.60–1.01). Stage 2 events were fifty versus ninety-three and stage 3 events thirty-four versus thirty-eight. Thus, the primary difference was concentrated in moderate injury. Among patients with duration data after moderate or severe injury, persistence beyond forty-eight hours was recorded in 32/82 versus 57/128; this conditional comparison had substantial uncertainty.

By ninety days, forty-one patients in each group had died, approximately 7%. Dialysis had been used in thirty versus thirty-five patients, 5.1% versus 5.9%, odds ratio 0.86 (95% CI 0.51–1.46). Major adverse kidney events occurred in 61/555 versus 60/564, 11.0% versus 10.6%, odds ratio 1.03 (95% CI 0.69–1.52). Intensive care and hospital stays were similar, as were changes in the urinary biomarker. No clear differences emerged in the prespecified adverse events, including arrhythmias, bleeding and unplanned return to theatre.^{2,3}

Critique

BigpAK-2 provides a persuasive randomised demonstration that a structured care strategy can reduce its specified early kidney-injury endpoint. Its size, multinational recruitment, masked primary-outcome adjudication and substantial differences in delivered care improve on the preceding small studies. The effect was not confined to an isolated biomarker change: conventional creatinine and urine-output definitions both contributed supportive findings.

The primary result still needs to be described at the correct clinical level. Preventing stage 2 or 3 acute kidney injury is valuable, but the trial did not demonstrate reductions in dialysis, death or persistent renal dysfunction. These outcomes were uncommon and their confidence intervals remain compatible with benefit as well as no benefit. Their

neutral point estimates do not erase the primary result, but they prevent the early endpoint from being presented as a proven improvement in long-term patient outcomes.

The distribution of injury severity sharpens this interpretation. Most of the numerical difference came from stage 2, with relatively similar stage 3 counts. Any-stage injury was not clearly reduced. One plausible interpretation is that the strategy limited progression to a more severe functional category in some patients rather than preventing every episode of kidney dysfunction. That explanation fits the pattern but is not a separately randomised mechanistic finding.

Urine output is particularly responsive to haemodynamic management, and oliguria accounted for much of the primary endpoint. A care strategy that changes perfusion, fluid decisions and vasoactive treatment can change urine output without necessarily reversing structural injury. This is not a reason to discard urine-output criteria, which identify clinically meaningful dysfunction. It is a reason to distinguish improved functional measurements from established prevention of lasting tissue damage. The creatinine-based reduction is reassuring, although smaller in absolute event numbers.

Fluid administration can also dilute creatinine, but the broadly similar cumulative fluid exposure makes a large dilution-only explanation less compelling. Conversely, unchanged stress-biomarker trajectories do not disprove a clinical effect: the marker may reflect a different stage or aspect of injury. Neither creatinine dilution nor biomarker behaviour provides a complete causal account. The trial established the effect of the strategy on its endpoint, not the precise pathway through which that effect arose.

The biomarker's role is frequently overstated in interpreting trials of this design. All randomised patients were biomarker-positive. There was no randomised comparison of biomarker-guided selection against clinical selection alone, and kidney outcomes were not systematically collected in biomarker-negative screened patients. The study therefore shows that the bundle worked in the selected population. It does not establish that testing was necessary, that this threshold was optimal, or that the same care would fail in clinically high-risk patients with a lower result.

Prognostic enrichment and treatment-effect prediction are different claims. A test can identify a high event rate without identifying patients who respond differently to treatment. BigpAK-2 used the first property to recruit an efficient study population. Determining the second would require a comparison across biomarker strata with an appropriate interaction analysis. Claims about cost-effectiveness also require the costs of testing, monitoring and treatment, not simply a higher control-event rate.

Only 15% of screened patients were randomised. That proportion reflects multiple eligibility conditions, including clinical risk, biomarker status, invasive access, postoperative location and consent. It should not be attributed solely to the biomarker. Patients with advanced chronic kidney disease, established acute kidney injury or urgent dialysis indications were excluded. These are important renal-risk groups, but their response to the intervention remains uncertain. The assumption that their risk is less modifiable was a design rationale, not a finding demonstrated by this trial.

The bundle comparison cannot determine which component caused benefit. Exploratory analyses associated avoiding hypotension and withholding renin–angiotensin system drugs with fewer primary events. Those comparisons were made among patients whose treatments and physiological responses had already been influenced by illness severity and clinical decisions. A patient in whom a target was easy to achieve may differ substantially from one in whom it was not. Adjustment does not restore the protection of randomisation for each individual component.

This is especially relevant to the glucose target. Full bundle adherence was achieved in fewer than half the intervention patients, partly because maintaining the specified glucose range was difficult. The study does not establish that tighter glucose control contributed benefit, nor that omitting it would preserve the entire effect. The intervention was assigned as a package and delivered with substantial flexibility. Its positive result should not override separate randomised evidence about glucose control or convert every component into an independently proven treatment.

Implementation itself was probably an active ingredient. Repeated assessment, education, reminders and a structured response to abnormalities can alter the timing and appropriateness of care even if total fluid or vasopressor doses change little. The recorded hypotension measure also combined an abnormal pressure with failure to intervene; it was not simply a blinded count of every episode below 65 mmHg. Describing it as a halving of all hypotension would overstate what was measured.

The control group is therefore central to transferability. Only 5% received every bundle element, with particularly low use of advanced monitoring and repeated passive leg raising. A unit already delivering these practices may obtain a smaller incremental effect. A unit with limited access to cardiac-output assessment, urinary testing or postoperative critical care may find the full strategy harder to reproduce. The accompanying editorial appropriately highlighted these practical questions and the unresolved value of alternative, simpler selection approaches.⁴

Open treatment allocation introduces the possibility of additional attention and co-interventions. Masked adjudication reduces direct outcome-assessment bias, but it cannot remove differences in bedside care beyond the explicitly listed measures. That is partly inherent in a pragmatic implementation strategy. It makes the result applicable to the organised care package, while limiting attribution to any particular device, fluid challenge or medication decision.

The adaptive design is a strength when its operating rules preserve error control and interim information is carefully contained. The supplementary methods describe staged disclosure of recruitment decisions without releasing treatment effects to investigators. The primary estimate and interval should be interpreted using the reported sequential method, rather than replaced with a convenient unadjusted calculation. The chapter reports these procedures from the publication and eligible supplementary methods; it does not claim an independent audit of the excluded protocol and statistical analysis plan.

Several reporting details deserve attention. The abstract and narrative refer to 1,176 available patients, while the outcome table and flowchart identify 1,172 with measured primary outcomes. Missingness was very small and is unlikely to overturn the result, but the denominators should remain explicit. Percentages for the stage distribution use patients with acute kidney injury as their denominator, and persistence analyses condition on having developed moderate or severe injury. Neither should be mistaken for the corresponding risk in all randomised patients.

Longer-term follow-up had more missing information, particularly for the composite requiring creatinine measurements. This adds uncertainty to the absence of a demonstrated downstream benefit. The trial was not designed to detect a modest change in a relatively uncommon ninety-day outcome. A larger or pooled evaluation could address that question, but association between early injury and later harm does not itself validate early injury as a surrogate through which every treatment must improve prognosis.

Finally, industry support and free testing matter to interpretation without invalidating the randomisation. The manufacturer's interest lies partly in wider biomarker adoption, whereas the randomised comparison isolates the care strategy within an already tested population. Keeping those claims separate is the most useful response to the commercial context. The evidence supports organised prevention in selected high-risk surgical patients; the incremental clinical and economic value of the screening assay remains a related, unanswered question.

Body of Evidence

The Sapphire observational validation programme established the predictive basis for this approach. In its final cohort of 728 critically ill adults, the combined urinary biomarkers predicted moderate or severe acute kidney injury within twelve hours with an area under the receiver-operating characteristic curve of 0.80. This was evidence of risk discrimination, not an intervention trial. It justified evaluating a response to the signal but could not establish what that response should be.¹

The single-centre PrevAKI trial tested a KDIGO-derived bundle after cardiac surgery in biomarker-positive patients. Any acute kidney injury occurred in 55.1% with the intervention and 71.7% with standard care, an absolute reduction of 16.6 percentage points (95% CI 5.5–27.9). Moderate or severe injury was also reduced, without clear effects on other secondary outcomes. Its relatively high event rate and single-centre setting left uncertainty about transportability.⁵

The original BigpAK trial studied 121 high-risk patients after major abdominal surgery. Overall acute kidney injury did not differ clearly between groups, although moderate or severe injury was reduced and a biomarker-range subgroup showed fewer episodes of any injury. Dialysis, hospital mortality and major kidney events at discharge did not clearly improve. The subgroup finding provided a hypothesis, while the small trial could not establish a dependable long-term clinical benefit.⁶

PrevAKI-Multicenter then focused primarily on implementation feasibility. Among 278 analysed patients, full adherence was 65.4% versus 4.2%. Any acute kidney injury was not reduced, but moderate or severe injury occurred in 14.0% versus 23.9%, an absolute difference of approximately ten percentage points. Because kidney injury was a secondary endpoint in a feasibility study, the result required confirmation. BigpAK-2 supplies a larger, more diverse test with moderate or severe injury as the primary outcome.⁷

The evidence for individual bundle elements is less uniform. NICE-SUGAR randomised 6,104 critically ill adults to a much tighter glucose target of 81–108 mg/dL or conventional control at or below 180 mg/dL. Intensive control increased ninety-day mortality, 27.5% versus 24.9%, and severe hypoglycaemia. Its target and population differ from BigpAK-2, so it is not a direct refutation of the latter's glucose range. It does show why the success of a broad bundle cannot be taken as general permission to intensify glycaemic control.⁸

Across the prevention trials, the most consistent signal is a reduction in moderate or severe postoperative acute kidney injury. Evidence for mortality, dialysis and lasting

renal protection remains less certain. BigpAK-2 makes structured supportive care harder to dismiss as merely an attractive principle, while leaving work to identify the most useful components, the simplest effective risk-selection strategy and the resources needed for reliable delivery.

Should high-risk surgical patients receive structured kidney-protective care?

BigpAK-2 supports a structured preventive strategy in patients matching its clinical and biomarker criteria. It reduced early moderate or severe acute kidney injury, without establishing longer-term benefit or the necessity of every bundle component.

References

1. Kashani K, Al-Khafaji A, Ardiles T, Artigas A, Bagshaw SM, Bell M, et al. Discovery and validation of cell cycle arrest biomarkers in human acute kidney injury. *Crit Care*. 2013;17(1):R25.
2. Zarbock A, Ostermann M, Forni L, Bode C, Wild L, Putensen C, et al. A preventive care strategy to reduce moderate or severe acute kidney injury after major surgery (BigpAK-2); a multinational, randomised clinical trial. *Lancet*. 2025;406(10521):2782-2791.
3. Zarbock A, Ostermann M, Forni L, Bode C, Wild L, Putensen C, et al. Supplementary methods and results to A preventive care strategy to reduce moderate or severe acute kidney injury after major surgery (BigpAK-2); a multinational, randomised clinical trial. *Lancet*. 2025;406(10521):2782-2791.
4. Siew ED. Bundled interventions to prevent acute kidney injury: finding the value in supportive care. *Lancet*. 2025;406(10521):2731-2733.
5. Meersch M, Schmidt C, Hoffmeier A, Van Aken H, Wempe C, Gerss J, et al. Prevention of cardiac surgery-associated AKI by implementing the KDIGO guidelines in high risk patients identified by biomarkers: the PrevAKI randomized controlled trial. *Intensive Care Med*. 2017;43(11):1551-1561.
6. Göcze I, Jauch D, Götz M, Kennedy P, Jung B, Zeman F, et al. Biomarker-guided Intervention to Prevent Acute Kidney Injury After Major Surgery: The Prospective Randomized BigpAK Study. *Ann Surg*. 2018;267(6):1013-1020.
7. Zarbock A, Küllmar M, Ostermann M, Lucchese G, Baig K, Cennamo A, et al. Prevention of Cardiac Surgery-Associated Acute Kidney Injury by Implementing the KDIGO Guidelines in High-Risk Patients Identified by Biomarkers: The PrevAKI-Multicenter Randomized Controlled Trial. *Anesth Analg*. 2021;133(2):292-302.
8. Finfer S, Chittock DR, Su SY, Blair D, Foster D, Dhingra V, et al. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med*. 2009;360(13):1283-1297.

ALBICS-AKI

Shehabi Y, Balachandran M, Al-Bassam W, Bailey M, Bellomo R, Bihari S, et al; ALBICS AKI Study Investigators. Postoperative 20% Albumin Infusion and Acute Kidney Injury in High-Risk Cardiac Surgery Patients: The ALBICS AKI Randomized Clinical Trial. JAMA Surg. 2025;160(8):835–844.

Introduction

Acute kidney injury after cardiac surgery reflects several interacting insults, including altered perfusion, inflammation, embolic injury and venous congestion. Patients with pre-existing renal impairment or complex operations are particularly vulnerable. An intervention that reliably reduced this complication would be valuable, but an attractive physiological mechanism is insufficient to establish renal protection. Changes in serum creatinine, dialysis requirements and sustained loss of kidney function also represent different levels of clinical consequence.¹

Albumin has several plausible protective properties. In addition to maintaining oncotic pressure, it carries substances, contributes antioxidant activity and may influence endothelial function. Low postoperative serum albumin is associated with adverse outcomes, but this association does not show that administering albumin reverses their cause. Hypoalbuminaemia can be a marker of inflammation, dilution, capillary leak and illness severity. Replacing a measured deficit is therefore not necessarily equivalent to treating the processes that produced it.

Concentrated albumin offers a smaller administered fluid volume than iso-oncotic preparations. Earlier studies demonstrated reductions in resuscitation volume or fluid balance, without establishing consistent improvement in major clinical outcomes.^{2,3} Separately, the PROTECTION trial found less creatinine-defined acute kidney injury with intravenous amino acids during cardiac surgery.⁴ Albumin also delivers protein, but its intravascular oncotic effects, metabolism and administration schedule differ substantially from a balanced amino-acid infusion. Benefit from one cannot be assumed for the other.

ALBICS-AKI tested a specific prophylactic strategy: adding a fixed postoperative infusion of 20% albumin to usual care in patients at increased renal risk after cardiopulmonary bypass. It was not a trial of albumin versus crystalloid for every resuscitation episode. Its result therefore informs whether this additional infusion should be prescribed to protect

the kidneys, while leaving other albumin indications and formulations as separate questions.¹

Synopsis

ALBICS-AKI was an investigator-initiated, open-label, parallel-group randomised trial at seven cardiac centres in Australia and Italy. Recruitment extended from July 2019 to August 2024, with a prolonged interruption around the COVID-19 pandemic; most participants were recruited from July 2022 onwards. Adults undergoing on-pump cardiac surgery were eligible if they had a preoperative estimated glomerular filtration rate below 60 mL/min/1.73 m², a combined coronary and valve procedure, multiple valve surgery or an operation involving the thoracic aorta.^{1,5}

Exclusions included very severe preoperative renal impairment, dialysis dependence, previous renal transplantation, serum albumin below 20 g/L, mechanical circulatory support specified by the trial and more than six hours in intensive care after the operation. Consequently, the trial selected patients at increased risk without including the full spectrum of advanced renal disease, profound hypoalbuminaemia or postoperative circulatory failure. Consent was obtained from patients or their surrogate decision makers.

A statistician generated a computerised allocation sequence using fixed-size permuted blocks, stratified by centre and preoperative renal function. Randomisation was delivered through a secure electronic platform. Patients assigned to albumin received 300 mL of 20% albumin over 15 hours, commencing within six hours of postoperative intensive care admission. This provided 60 g of albumin at 4 g per hour. In controls, 20% albumin was prohibited during the first 24 hours. Both groups otherwise received usual care, including crystalloids and, when clinicians chose, 4% or 5% albumin.

Haemodynamic targets, vasopressors, inotropes, diuretics, transfusion decisions and kidney replacement therapy were left to treating clinicians. This made the comparison pragmatic but allowed management to respond to knowledge of treatment allocation. It also meant that the contrast was additional scheduled concentrated albumin versus usual postoperative fluid practice, rather than exposure to any albumin versus none.

The primary outcome was acute kidney injury of stages 1–3 using the creatinine-based Kidney Disease: Improving Global Outcomes definition. Daily serum creatinine was collected while patients remained in hospital; urine-output criteria were not used to define the primary endpoint. Secondary outcomes included major adverse kidney events at day 28 or hospital discharge, mortality, more severe acute kidney injury, dialysis,

lengths of stay, ventilation duration and days free of vasoactive treatment. The major adverse kidney event composite combined death, dialysis dependence and sustained major creatinine elevation. Safety outcomes included significant arrhythmia, clinically diagnosed fluid overload, transfusion, reintubation and intensive care readmission.

The planned sample of 620 allowed for approximately 5% attrition and provided 80% power to detect a reduction in acute kidney injury from 30% to 20%. No interim analysis was planned. The published analyses included unadjusted comparisons and generalised linear mixed models estimating relative risks after accounting for site and baseline renal-function stratum. Site was modelled as a random effect. Duration outcomes used competing-risk methods to account for death. Secondary comparisons were not adjusted for multiplicity.

Of 954 screened patients, 621 were randomised. Three subsequently withdrew consent and seven were judged ineligible, leaving a modified intention-to-treat population of 611: 307 assigned albumin and 304 usual care. Primary outcome data were complete in this population. Mean age was approximately 69 years, nearly three-quarters were men, 46% had an estimated glomerular filtration rate below 60 mL/min/1.73 m² and almost 89% underwent elective surgery. Around half had a combined operation; the median EuroSCORE II was approximately 3.2. Allocation produced broadly similar groups, although regular insulin treatment was more frequent among controls.

Treatment separation was substantial. The median intervention dose was the planned 300 mL, and adherence was approximately 98%. Usual-care patients received iso-oncotic albumin more frequently by the end of day 2: 71.4% versus 54.4%. Median amounts across the groups were 500 versus 250 mL. Thus, other albumin use partly differed in response to the allocated strategy. Total intravenous fluid administration was similar, at approximately 2.5 litres, and fluid balance was negative in both groups by the end of the second postoperative day. Serum albumin at 24 hours was higher with the intervention, averaging 37.4 versus 32.1 g/L.

Acute kidney injury occurred in 150 of 307 albumin patients (48.9%) and 132 of 304 controls (43.4%). The reported absolute difference was 5.4 percentage points (95% confidence interval -2.5 to 13.3). The unadjusted relative risk was 1.13 (0.95–1.34; P=0.18). After adjustment for the randomisation strata, the relative risk was 1.12 with a considerably narrower confidence interval of 1.04–1.21 (P=0.003). Both estimates pointed towards more kidney injury with albumin, but only the main adjusted model excluded no difference.¹

Among patients with baseline estimated glomerular filtration below 60 mL/min/1.73 m², acute kidney injury occurred in 66.2% versus 57.6%. The corresponding adjusted relative risk was 1.14 (95% confidence interval 1.07–1.22). The report did not establish significant heterogeneity of treatment effect across its prespecified subgroups. Accordingly, this numerically larger absolute difference should not be interpreted as proof that the effect was confined to patients with chronic kidney disease.

Major adverse kidney events occurred in 33 patients (10.7%) versus 29 (9.5%). Dialytic therapy was used in 10 (3.3%) versus seven (2.3%), and six versus two patients died by day 28 or hospital discharge. Severe renal outcomes and recovery-duration measures did not show a persuasive clinical advantage with albumin. Red-cell transfusion was more frequent, in 116 patients (37.8%) versus 91 (29.9%; $P=0.04$). Other reported safety comparisons did not provide a clear difference, although the trial had limited power for uncommon events.

The ordinary supplement included a post hoc analysis treating site as a fixed effect. Because of convergence difficulties, the smallest site was combined with another; the adjusted relative risk was 1.14 (95% confidence interval 0.99–1.31). An adjusted logistic analysis gave an odds ratio of 1.29 (0.92–1.81). These analyses maintained the adverse direction of the point estimate while demonstrating that its statistical significance depended on model specification.⁵

Critique

The most secure clinical conclusion is that this scheduled albumin infusion did not provide the expected renal protection. The observed effect ran in the opposite direction, albumin concentrations increased as intended, and the planned treatment was usually delivered. Failure to administer enough study treatment is therefore an unconvincing explanation for the absence of benefit. Whether the trial establishes a precisely quantified harmful effect requires more caution because the strength of the statistical evidence varied substantially between analyses.

The difference between adjusted and unadjusted results is unusually important. Adjustment for prognostic stratification factors can improve precision and is methodologically reasonable. Here, however, the point estimate changed very little while its confidence interval narrowed markedly. There were only seven centres, with highly unequal recruitment and substantial variation in acute kidney injury rates. The supplemental fixed-site model and adjusted logistic model did not reproduce the clear exclusion of the null found in the principal random-effects relative-risk model. These are not interchangeable estimands, and none automatically invalidates the others. Together,

they make the level of certainty about harm less robust than the abstract's single adjusted interval might suggest.

The supplement explicitly acknowledges the large reduction in standard error and the small number of sites. This transparency is useful. It also means the alternative estimates should not be omitted from a critical appraisal. Without patient-level data and independent model diagnostics, it is not possible to determine whether the narrow interval adequately reflects all relevant uncertainty. The appropriate response is to present the consistent direction and the dependence of statistical significance on the model, rather than either declaring the result wholly neutral or treating the adjusted P value as definitive proof of the magnitude of harm.⁵

The selected primary endpoint is clinically relevant but not synonymous with permanent kidney damage. Creatinine is an objective measurement, reducing susceptibility to subjective outcome classification. Nevertheless, it is influenced by filtration, creatinine generation and distribution volume. Albumin may alter intravascular volume and filtration physiology without every creatinine change representing structural tubular injury. Conversely, excluding urine-output criteria may miss episodes of acute kidney injury. The trial cannot establish which biological mechanism produced its creatinine signal because it did not systematically measure structural injury, renal perfusion or venous congestion.

This distinction should not be used to dismiss stage 1 acute kidney injury as harmless. It is associated with later risk and can complicate care. However, the absence of a clear reduction in dialysis, major adverse kidney events or length of stay means there is no corroborating patient-centred benefit to balance against the primary adverse direction. Only 17 patients received dialysis and eight died. These small totals make estimates of effects on the most serious outcomes unstable. One adjusted mortality interval in the main table excludes unity, whereas unadjusted and supplementary logistic intervals are much wider and include it. With sparse events and multiple secondary analyses, mortality harm is not established.

Open treatment allocation may have influenced co-interventions. Clinicians used more iso-oncotic albumin in controls, potentially reducing the contrast in total albumin exposure or reflecting different perceptions of volume needs. This is part of the treatment-policy effect of introducing a fixed concentrated infusion into usual practice. It does not mean the randomised comparison was invalid. It does mean that the trial cannot isolate a pure molecular effect of 60 g of albumin while holding all subsequent fluid and haemodynamic management constant.

The sensitivity analysis excluding controls who received substantial iso-oncotic albumin is less reliable for causal inference. It selects patients according to a post-randomisation clinical decision in one group only. Those requiring more resuscitation may differ in illness severity and risk of kidney injury. Weighting by measured baseline characteristics can address some imbalance, but cannot guarantee removal of confounding by events that developed after surgery. The weighted estimate remained adverse in direction but imprecise. This analysis should not be treated as a second randomised albumin-versus-no-albumin trial.⁵

The red-cell transfusion finding deserves attention without overstatement. The absolute difference was approximately eight percentage points, but transfusion was one of several safety and tertiary comparisons, with no multiplicity adjustment. Thresholds were left to clinicians, and masking was absent. Dilution, altered haemostasis, bleeding and local decision-making could all contribute. Similar fluid balance and no clear excess of return to theatre make a simple explanation based on major bleeding insufficient. The trial identifies a possible additional disadvantage rather than proving a specific haemostatic mechanism.

The population defines an important boundary. Patients with serum albumin below 20 g/L, dialysis dependence, previous transplantation and selected forms of mechanical support were excluded. Most procedures were elective. The study therefore does not directly answer whether an individual patient with profound hypoalbuminaemia and a separate accepted indication should receive albumin, or how best to resuscitate severe postoperative shock. Equally, extrapolating benefit to these unstudied patients would be unwarranted. Within the population studied, a routine prophylactic infusion lacks an established clinical rationale after this result.

Later correspondence proposed renal venous congestion as a mechanism for the adverse renal signal. Another letter discussed possible acid-base effects of concentrated albumin. The investigators replied that central venous pressure and renal Doppler measures were not recorded, and that the net acid-base effect depends on the preparation's constituents and the wider clinical response. These exchanges are physiologically informative but do not supply missing mechanistic measurements.^{6,7,8} The chapter's interpretation therefore does not attribute the outcome to congestion, acidosis or albuminuria as established causes.

The trial received academic and hospital support, with seed funding from CSL Behring. The report states that funders did not participate in design, conduct, data analysis or publication decisions, while the manufacturer and academic sponsor participated in pharmacovigilance. The investigators' willingness to report an unfavourable result is

valuable. The trial's principal contribution is a direct challenge to prophylactic use supported by physiological reasoning rather than demonstrated patient benefit.

Body of Evidence

Albumin trials differ in concentration, timing, indication and comparator. The SAFE trial randomised 6,997 intensive care patients to 4% albumin or saline for resuscitation. Overall 28-day mortality was similar, with a relative risk of 0.99 (95% confidence interval 0.91–1.09). SAFE established that these resuscitation strategies had similar overall outcomes in a heterogeneous intensive care population. It did not test concentrated postoperative albumin as a nephroprotective drug and should not be used to assume equivalence for that distinct intervention.⁹

The 2022 ALBICS trial compared 4% albumin with Ringer acetate for cardiopulmonary bypass priming and perioperative volume replacement. Among 1,386 patients completing the trial, a major adverse event occurred in 37.1% versus 33.8%, respectively, with a relative risk of 1.10 (95% confidence interval 0.95–1.27). Albumin reduced the volume of study fluid administered but did not improve the principal clinical composite. This provides an important precedent: a fluid-sparing effect does not itself establish better postoperative outcomes.¹⁰

SWIPE compared 20% with 4–5% albumin in 321 intensive care patients requiring resuscitation. Concentrated albumin reduced resuscitation volume over 48 hours by approximately 600 mL and produced a lower cumulative fluid balance. The study supported further evaluation but was not a definitive trial of uncommon renal or mortality outcomes. Its comparator was another albumin concentration, and patients received fluid because resuscitation was considered necessary. These differences limit direct extrapolation to a fixed prophylactic infusion after cardiac surgery.²

HAS FLAIR-II moved closer to the postoperative setting. Of 480 randomised cardiac surgery patients, 466 contributed to the primary analysis. A 20% albumin bolus strategy reduced positive fluid balance compared with a crystalloid-based strategy, but did not significantly shorten the cumulative duration of vasopressor therapy. Median durations were seven and 10.8 hours, with a difference of –3.8 hours and a confidence interval extending from –8 to 0.4 hours. The population had a lower median EuroSCORE II than ALBICS-AKI, and fluid was administered in response to a perceived need for a bolus.³

An earlier single-centre trial by Lee and colleagues studied 220 patients with serum albumin below 4.0 g/dL undergoing off-pump coronary bypass surgery. Preoperative 20% albumin, with dose determined by baseline concentration, was associated with less

creatinine-defined acute kidney injury: 13.7% versus 25.7% among analysed patients. Severe kidney injury and other major outcomes did not differ clearly. The timing, off-pump operation, selected albumin concentration and smaller sample distinguish this result from ALBICS-AKI. It cannot establish that the later postoperative on-pump strategy should have the same effect.¹¹

PROTECTION tested an explicitly different intervention: a balanced intravenous amino-acid mixture during cardiac surgery and for up to three days. In 3,511 patients, acute kidney injury occurred in 26.9% with amino acids and 31.7% with placebo, with a relative risk of 0.85 (95% confidence interval 0.77–0.94). Stage 3 injury was also less frequent. These findings support the possibility of recruiting renal functional reserve, but do not establish a class effect of all protein-containing infusions. Albumin's oncotic action and delivery as an intact protein make substitution on physiological grounds inappropriate.⁴

Taken together, the evidence available by the end of 2025 does not support routine albumin administration simply to improve renal outcomes after cardiac surgery. The most consistent advantages of concentrated albumin concern administered volume and fluid balance in selected resuscitation settings. ALBICS-AKI addressed a separate prophylactic indication and found no benefit, with a concerning adverse renal direction whose statistical certainty depends on the analysis. This is sufficient to argue against adding the tested infusion to usual care, without extending the result to every formulation or established indication for albumin.

Should a routine postoperative infusion of 20% albumin be used to prevent acute kidney injury after high-risk cardiac surgery?

No. ALBICS-AKI showed no renal protection and a signal towards more acute kidney injury, although the statistical certainty of harm varied between models.

References

1. Shehabi Y, Balachandran M, Al-Bassam W, Bailey M, Bellomo R, Bihari S, et al; ALBICS AKI Study Investigators. Postoperative 20% Albumin Infusion and Acute Kidney Injury in High-Risk Cardiac Surgery Patients: The ALBICS AKI Randomized Clinical Trial. *JAMA Surg.* 2025;160(8):835-844.
2. Mårtensson J, Bihari S, Bannard-Smith J, Glassford NJ, Lloyd-Donald P, Cioccarl L, et al. Small volume resuscitation with 20% albumin in intensive care: physiological effects : The SWIPE randomised clinical trial. *Intensive Care Med.* 2018;44(11):1797-1806.

3. Wigmore GJ, Deane AM, Presneill JJ, Eastwood G, Serpa Neto A, Maiden MJ, et al. Twenty percent human albumin solution fluid bolus administration therapy in patients after cardiac surgery-II: a multicentre randomised controlled trial. *Intensive Care Med.* 2024;50(7):1075-1085.
4. Landoni G, Monaco F, Ti LK, Baiardo Redaelli M, Bradic N, Comis M, et al. A Randomized Trial of Intravenous Amino Acids for Kidney Protection. *N Engl J Med.* 2024;391(8):687-698.
5. Shehabi Y, et al; ALBICS AKI Study Investigators. Supplemental online content to the ALBICS AKI randomized clinical trial. *JAMA Surg.* 2025.
6. Beaubien-Souligny W, Deschamps J, Denault A. Albumin Infusion and Kidney Injury in Cardiac Surgery. *JAMA Surg.* 2025;160(12):1395.
7. Saro R, Orso D, Bove T. Albumin Infusion and Kidney Injury in Cardiac Surgery. *JAMA Surg.* 2025;160(12):1395-1396.
8. Balachandran M, Pakavakis A, Shehabi Y. Albumin Infusion and Kidney Injury in Cardiac Surgery-Reply. *JAMA Surg.* 2025;160(12):1396-1397.
9. Finfer S, Bellomo R, Boyce N, French J, Myburgh J, Norton R. A comparison of albumin and saline for fluid resuscitation in the intensive care unit. *N Engl J Med.* 2004;350(22):2247-2256.
10. Pesonen E, Vlasov H, Suojaranta R, Hiippala S, Schramko A, Wilkman E, et al. Effect of 4% Albumin Solution vs Ringer Acetate on Major Adverse Events in Patients Undergoing Cardiac Surgery With Cardiopulmonary Bypass: A Randomized Clinical Trial. *JAMA.* 2022;328(3):251-258.
11. Lee EH, Kim WJ, Kim JY, Chin JH, Choi DK, Sim JY, et al. Effect of Exogenous Albumin on the Incidence of Postoperative Acute Kidney Injury in Patients Undergoing Off-pump Coronary Artery Bypass Surgery with a Preoperative Albumin Level of Less Than 4.0 g/dl. *Anesthesiology.* 2016;124(5):1001-1011.



Haematology Trials

FARES-II

Karkouti K, Callum JL, Bartoszko J, Tanaka KA, Knaub S, Brar S, et al. Prothrombin Complex Concentrate vs Frozen Plasma for Coagulopathic Bleeding in Cardiac Surgery: The FARES-II Multicenter Randomized Clinical Trial. JAMA. 2025;333(20):1781-1792.

Introduction

Bleeding after cardiopulmonary bypass can reflect a surgical source, platelet dysfunction, reduced fibrinogen, residual anticoagulation or depletion of coagulation factors. These problems overlap, but they do not require identical treatment. When impaired thrombin generation from factor deficiency is an important contributor, frozen plasma offers a broad replacement product. Its administration also introduces considerable volume and requires preparation and compatibility procedures during a time-critical episode.

Four-factor prothrombin complex concentrate supplies factors II, VII, IX and X in a small volume. It can be prepared and administered rapidly, but does not replace the full range of proteins present in plasma. A stronger procoagulant effect could improve haemostasis while also increasing thrombosis. The relevant question is therefore more demanding than whether the international normalised ratio falls: does targeted replacement control bleeding, reduce additional treatment and preserve safety?

Small cardiac surgical trials suggested that prothrombin complex concentrate might substitute for plasma, with possible reductions in transfusion. FARES-II was the larger confirmatory comparison. It evaluated a particular four-factor product and dosing strategy in patients with bleeding after bypass who required coagulation factor replacement, rather than routine prophylactic administration to every patient undergoing cardiac surgery.¹

Synopsis

FARES-II was a phase 3, randomised noninferiority trial at ten Canadian and two United States hospitals. Patients were recruited from November 2022 to May 2024 and followed to postoperative day 30. Octapharma and the Canadian Institutes of Health Research provided funding. The manufacturer participated in final design, conduct, interpretation and manuscript review; data management and analysis were outsourced, and an independent monitoring committee reviewed safety. Treating clinicians became unblinded when treatment began, while patients and outcome assessors remained masked.¹

Adults undergoing cardiac surgery with cardiopulmonary bypass were considered when clinicians ordered plasma or prothrombin complex concentrate for actual or anticipated bleeding after bypass. Important exclusions included heart transplantation, ventricular assist device procedures, thoracoabdominal aneurysm repair, severe right heart failure, very high immediate mortality risk, recent thromboembolism and recent or intraoperative use of concentrate to reverse oral anticoagulation. The study was not a trial of anticoagulant reversal.²

Randomisation occurred when the blood bank or pharmacy received the order. A site-stratified, permuted-block schedule was held in sequentially numbered opaque envelopes. Products were transported in weighted, tamper-proof boxes, preserving concealment until the clinical team made its final decision to administer treatment. Administration ordinarily required at least moderate bleeding and a point-of-care international normalised ratio of at least 1.5 after heparin neutralisation. The laboratory requirement could be waived when severe bleeding demanded immediate treatment.

This distinction between ordering and administering treatment explains much of the subsequent attrition. Some patients randomised after an anticipated need did not ultimately meet the administration criteria. United States participants consented before surgery, whereas Canadian sites used an approved delayed-consent process after surgery. Some treated patients subsequently declined participation or could not provide consent, so the primary population was restricted to randomised, treated and consenting patients.

The concentrate group received Octaplex or Balfaxar, 1,500 IU at body weight 60 kg or less and 2,000 IU above 60 kg. The plasma group received three or four units using the same weight threshold. One further assigned dose was permitted within twenty-four hours. Additional factor replacement ordinarily used plasma in both groups, although rescue commercial concentrate was permitted after two assigned plasma doses if

clinicians judged it necessary. The intended maximum cumulative concentrate exposure was approximately 50 IU/kg.¹

Other treatment followed established local bleeding-management practice rather than a single trial-wide transfusion algorithm. Antifibrinolytics were routinely used. Viscoelastic testing could supplement point-of-care international normalised ratio assessment. Decisions about red cells, platelets, fibrinogen replacement and surgical re-exploration remained with the clinical teams.

The primary outcome was an effective haemostatic response, defined by the absence of further haemostatic interventions from sixty minutes to twenty-four hours after the first assigned dose began. Failure included surgical reopening for bleeding, another assigned dose, non-red-cell allogeneic transfusion or a coagulation factor concentrate. Red-cell transfusion alone did not constitute primary failure. The first hour was excluded to allow administration of the assigned product.

Secondary outcomes included a global haemostatic response incorporating haemoglobin decline, severe or massive bleeding, additional transfusions, chest-drain output and change in international normalised ratio. Death, thromboembolic events and other adverse outcomes were assessed through thirty days. Several transfusion summaries used different starting points: surgery, the end of bypass or treatment initiation. These windows should not be interchanged.^{1,2}

A target of 410 evaluable patients provided at least 90% power for noninferiority, assuming response rates of 70% with concentrate and 65% with plasma. The noninferiority margin was ten percentage points, with one-sided alpha 0.025. Superiority testing was planned if noninferiority was established. An administrative unblinded interim analysis after approximately 200 enrolments allowed sample-size reassessment or futility review; no change was recommended. Secondary analyses, including individual primary components, were exploratory and unadjusted for multiplicity.

Of 538 patients enrolled, 528 were randomised and 450 received treatment. Thirty treated patients lacked consent for the primary analysis, leaving 420: 213 assigned concentrate and 207 plasma. Overall, 108 randomised patients, 20.4%, were absent from the primary population, although most had not received treatment. Eleven further patients were excluded from the per-protocol analysis. One hospital contributed 158 participants, with a median of approximately twenty-two per site.

Median age was 66 years, 74% were male, and 70% underwent procedures classified as complex. Approximately one-fifth had nonelective surgery. Mean bypass duration was

nearly three hours. First-dose concentrate exposure averaged 23.9 IU/kg, while plasma averaged 11.8 mL/kg. Median time from bypass completion to starting treatment was 41 versus 45 minutes. Administration itself was faster with concentrate, seven versus twenty-six minutes. A second assigned dose was given to 17.4% versus 22.7%; some doses occurred within the first hour and therefore did not count towards primary failure.¹

Effective haemostasis occurred in 166/213 concentrate patients, 77.9%, and 125/207 plasma patients, 60.4%. The absolute difference was 17.6 percentage points (95% CI 8.7–26.4), satisfying both noninferiority and superiority, $P < 0.001$. Equivalently, treatment failure occurred in 22.1% versus 39.6%, relative risk 0.56 (95% CI 0.41–0.75). Results were similar in the per-protocol population. Further platelet treatment and another assigned dose after the first hour contributed substantially to the composite difference.

Red-cell transfusion within twenty-four hours of treatment initiation occurred in 45.5% versus 63.8%, relative risk 0.71 (95% CI 0.60–0.85). Over twenty-four hours after bypass, the modelled mean number of red-cell, platelet and non-investigational plasma units was 6.6 versus 9.3, a difference of 2.7 units (95% CI 1.0–4.4). Including assigned plasma increased the apparent difference because plasma was itself the control treatment. Platelet doses were converted to four units for these summaries.

Severe or massive bleeding occurred in 14.1% versus 27.5%, relative risk 0.51 (95% CI 0.34–0.76). Modelled twenty-four-hour chest-drain loss was 691 versus 923 mL, a difference of 232 mL (95% CI 126–338). Reopening for bleeding occurred in eleven versus fifteen patients, with an imprecise relative risk of 0.71 (95% CI 0.34–1.50). Thus, several bleeding measures favoured concentrate, while reoperation alone did not establish a difference.¹

Serious adverse events affected 77 patients, 36.2%, versus 98, 47.3%, relative risk 0.76 (95% CI 0.61–0.96). Acute kidney injury occurred in 22, 10.3%, versus 39, 18.8%, relative risk 0.55 (95% CI 0.34–0.89). Thromboembolic events affected eighteen versus fifteen patients, 8.5% versus 7.2%, including five strokes in each group. Seven versus eight patients died. Ventilation duration, intensive care stay and hospital stay were similar. These safety and recovery outcomes were exploratory, not separately powered confirmatory endpoints.

Critique

FARES-II offers a clinically relevant comparison at the point when factor replacement is already considered necessary. Its treatment effect on the primary endpoint was substantial, and the superiority test followed a planned sequence rather than an opportunistic switch after an inconclusive noninferiority result. Concordant reductions in red-cell use and drain output make it less likely that the entire result arose from the construction of the composite alone.

The primary endpoint nevertheless remains sensitive to unblinded clinical decisions. A clinician who knows that concentrate has been given may respond differently to continued oozing than one who has given plasma. Decisions to administer platelets, fibrinogen or rescue factor replacement are influenced by thresholds, expectations and local practice. Masked outcome assessors can accurately record those interventions, but cannot remove bias from the decisions that generated them.

This is a limitation of interpretation, not a reason to dismiss a useful pragmatic outcome. Avoiding additional haemostatic treatment is relevant to patients and blood services. The concern is how much of the effect reflects superior physiological haemostasis and how much reflects the clinical response to a known product. Lower measured drain output provides an additional observation less directly determined by treatment preference, although it remains influenced by the broader perioperative course.

The composite also combines interventions of unequal importance. Avoiding platelet transfusion or a repeat dose matters, but it is not equivalent to avoiding reoperation. The component analysis shows that the result should not be advertised as a demonstrated reduction in surgical reopening. Conversely, a trial of 420 patients has limited power for an uncommon individual component. The correct conclusion is stronger overall haemostatic effectiveness, with uncertainty about some of its most consequential individual outcomes.

Timing affects the endpoint. Concentrate finished approximately nineteen minutes sooner, and any treatment during the first sixty minutes was excluded from primary failure. Faster administration is a real advantage of the product, but the outcome evaluates the entire delivery strategy rather than an isolated pharmacological comparison at equal factor exposure and equal completion time. The total number of second doses was larger than the number counted as primary failures because some were given during that initial window.

The noninferiority margin would have permitted a ten-point reduction in effective haemostasis, a potentially important loss if noninferiority were the only finding. Here,

the observed confidence interval favoured concentrate and excluded no difference, making the margin less decisive to the practical conclusion. It remains inaccurate to describe the trial as a safety noninferiority study: the formal margin applied to haemostatic effectiveness, not thrombosis or mortality.

The analysis population is another central issue. Excluding untreated patients whose final eligibility assessment preceded opening the concealed product is more defensible than excluding patients after an observed response. Nonetheless, the primary analysis did not retain everyone randomised. Thirty treated patients were omitted because consent was unavailable, and delayed consent can relate to postoperative circumstances. Similar exclusion proportions and anticipated attrition do not prove that selection was harmless.

The concealment procedures reduce the opportunity for allocation to influence the decision to treat, and the per-protocol analysis was consistent. These are reassuring features. They do not recreate a complete intention-to-treat comparison, particularly for consent-related exclusions after exposure. The trial's estimate is most securely understood in the treated, consenting population that was analysed. Transparency about that population is preferable to calling all 420 patients the full randomised cohort.

Transfusion results require especially careful arithmetic. When the control intervention consists of several plasma units, including those units in total transfusion creates an automatic difference even if subsequent bleeding is identical. The comparison excluding assigned plasma is therefore more informative about additional treatment: 6.6 versus 9.3 modelled units within twenty-four hours after bypass. It still combines different products, with each platelet dose represented as four units. This is a useful resource measure, not a direct statement that the average patient avoided 2.7 red-cell units.

The red-cell-specific findings are more readily interpretable and support an early benefit. Fewer concentrate patients received red cells during the first day after treatment, and the amount transfused was lower. By seven days, the proportion exposed to any red cells was more similar, although total quantity still favoured concentrate. The intervention may reduce early treatment burden without ensuring complete avoidance of transfusion over the entire admission.

The kidney-injury and serious-adverse-event findings are encouraging but less definitive. They came from multiple exploratory comparisons without adjustment, and the study was not powered around renal injury. Less bleeding, less plasma volume and fewer transfusions offer plausible pathways, but no one mechanism was established. Baseline characteristics were generally similar, although five plasma-group patients and no

concentrate-group patients were already receiving dialysis. The renal result warrants confirmation rather than a claim that concentrate is an established kidney-protective treatment.

Thromboembolism remains the most important unresolved safety boundary. Events were numerically slightly more common with concentrate, though the sample was too small to resolve a modest difference. Five strokes in each group are reassuring at face value but provide limited precision. A nonsignificant comparison does not establish equal thrombotic safety. The trial also excluded recent thromboembolism and several high-risk surgical groups, so those results cannot simply be extended to them.

Safety ascertainment used treatment-emergent events through thirty days and broad thromboembolic categories. The serious-adverse-event table additionally retained limited information from two patients without full consent under ethics approval. Such details matter when comparing denominators or interpreting the total number of events. Pharmacovigilance rates cited elsewhere are not interchangeable with actively collected postoperative event rates in a population already at substantial thrombotic risk.

The tested dose and formulation matter. FARES-II used a particular four-factor preparation at an initial exposure averaging about 24 IU/kg, with one repeat dose permitted. It does not establish equivalence among all concentrates, an optimal dose, or the safety of unrestricted repeated dosing. Nor does it show that replacing factors II, VII, IX and X can correct every cause of post-bypass coagulopathy. Low fibrinogen, platelet dysfunction, residual heparin and surgical bleeding still require appropriate assessment and treatment.

Point-of-care assessment helps define the population. Most patients needed documented international normalised ratio elevation in addition to bleeding, while severe bleeding allowed an exception. The report does not provide a clear proportion treated under that exception. This limits interpretation of how much benefit depended on laboratory confirmation. It should not be converted into support for empirical concentrate in any bleeding patient, particularly where factor deficiency has not been assessed.

Generalisability is further shaped by experienced cardiac centres, routine antifibrinolytics and established local transfusion pathways. One hospital supplied a large share of the cohort. The broad inclusion of complex operations is valuable, but transplant, ventricular assist device and some very high-risk patients remained outside

the comparison. The accompanying editorial highlighted both the efficacy signal and the practical importance of testing access, acquisition cost and thrombotic selection.³

Finally, the financial claim requires a full evaluation. Concentrate reduces plasma use by substitution, but it is itself derived from pooled plasma and has substantial acquisition costs. Reduced additional transfusions may offset part of that cost; preparation time, wastage, staffing and complications also matter. FARES-II did not provide a definitive cost-effectiveness analysis. Manufacturer participation in design and interpretation makes it particularly useful to keep demonstrated clinical results separate from projections about national blood supply or hospital savings.

Body of Evidence

The preceding FARES pilot analysed 101 treated, consenting patients at two Canadian hospitals using the same weight-banded doses. Effective haemostasis at twenty-four hours occurred in 76% with concentrate and 66% with plasma, an imprecise difference. Total transfusion favoured concentrate, but the comparison weakened when assigned plasma was removed. The pilot established feasibility and suggested benefit; FARES-II supplied the larger confirmatory test.⁴

PROPHECY randomised fifty bleeding cardiac surgical patients to concentrate at 15 IU/kg or plasma at 15 mL/kg. It was principally a feasibility study and demonstrated different immediate factor-replacement profiles: plasma increased several proteins absent from concentrate, while concentrate produced higher factor II and X levels. Its size did not resolve important clinical benefits or uncommon harms. It also used a lower concentrate dose than FARES-II.⁵

Smith and colleagues subsequently studied one hundred treated patients from a larger pre-randomised surgical cohort. Concentrate at 15 IU/kg did not clearly reduce the primary chest-drain output measure compared with plasma. It improved coagulation tests and reduced post-treatment intraoperative red-cell exposure, without clear differences in broader postoperative outcomes. These findings are not necessarily inconsistent with FARES-II, given differences in dose, power, selection and endpoint. They do not prove that dose alone explains the contrasting primary results.⁶

The TACS stepped-wedge trial addressed the surrounding system of care. Across 7,402 cardiac surgical patients, introducing point-of-care testing within an integrated transfusion algorithm reduced red-cell transfusion, adjusted relative risk 0.91 (95% CI 0.85–0.98), and reduced platelet use and major bleeding. It did not establish a reduction in major complications. This supports the importance of an organised diagnostic and

treatment pathway, but does not independently establish the superiority of one factor-replacement product.⁷

PROCOAG provides an important caution from another setting. In 324 analysed trauma patients at risk of massive transfusion, adding 25 IU/kg four-factor concentrate to a strategy already using early ratio-based blood products did not reduce twenty-four-hour product consumption. Thromboembolic events occurred in 35% versus 24%, relative risk 1.48 (95% CI 1.04–2.10). This was empirical addition in trauma, not replacement of plasma for selected post-bypass factor deficiency. The difference in indication explains why the studies should not be pooled into a universal claim that concentrate is either beneficial or harmful whenever bleeding occurs.⁸

FARES-II materially strengthens the case for concentrate as an alternative to plasma in the population and dosing range studied. Its most secure contribution is better short-term haemostatic effectiveness and reduced additional transfusion. Mortality benefit, renal protection, comparative cost and safety in excluded high-risk groups remain less certain. Adoption is best understood as a change within a structured bleeding-management pathway, with continued attention to the specific deficiency being treated.

Should four-factor prothrombin complex concentrate replace plasma for selected post-bypass coagulopathic bleeding?

FARES-II supports preferential use at the studied doses when coagulation factor replacement is indicated. Better haemostasis and less transfusion are established more clearly than renal benefit or the absence of thrombotic harm.

References

1. Karkouti K, Callum JL, Bartoszko J, Tanaka KA, Knaub S, Brar S, et al. Prothrombin Complex Concentrate vs Frozen Plasma for Coagulopathic Bleeding in Cardiac Surgery: The FARES-II Multicenter Randomized Clinical Trial. JAMA. 2025;333(20):1781-1792.
2. Karkouti K, Callum JL, Bartoszko J, Tanaka KA, Knaub S, Brar S, et al. Supplementary methods and results to Prothrombin Complex Concentrate vs Frozen Plasma for Coagulopathic Bleeding in Cardiac Surgery: The FARES-II Multicenter Randomized Clinical Trial. JAMA. 2025;333(20):1781-1792.

3. Wang R, Bennett-Guerrero E. Is It Time to Replace Plasma With Prothrombin Complex Concentrate in Cardiac Surgery? *JAMA*. 2025;333(20):1777-1778.
4. Karkouti K, Bartoszko J, Grewal D, Bingley C, Armali C, Carroll J, et al. Comparison of 4-Factor Prothrombin Complex Concentrate With Frozen Plasma for Management of Hemorrhage During and After Cardiac Surgery: A Randomized Pilot Trial. *JAMA Netw Open*. 2021;4(4):e213936.
5. Green L, Roberts N, Cooper J, Agarwal S, Brunskill SJ, Chang I, et al. Prothrombin complex concentrate vs. fresh frozen plasma in adult patients undergoing heart surgery - a pilot randomised controlled trial (PROPHECY trial). *Anaesthesia*. 2021;76(7):892-901.
6. Smith MM, Schroeder DR, Nelson JA, Mauermann WJ, Welsby IJ, Pochettino A, et al. Prothrombin Complex Concentrate vs Plasma for Post-Cardiopulmonary Bypass Coagulopathy and Bleeding: A Randomized Clinical Trial. *JAMA Surg*. 2022;157(9):757-764.
7. Karkouti K, Callum J, Wijeyesundera DN, Rao V, Crowther M, Grocott HP, et al. Point-of-Care Hemostatic Testing in Cardiac Surgery: A Stepped-Wedge Clustered Randomized Controlled Trial. *Circulation*. 2016;134(16):1152-1162.
8. Bouzat P, Charbit J, Abback PS, Huet-Garrigue D, Delhaye N, Leone M, et al. Efficacy and Safety of Early Administration of 4-Factor Prothrombin Complex Concentrate in Patients With Trauma at Risk of Massive Transfusion: The PROCOAG Randomized Clinical Trial. *JAMA*. 2023;329(16):1367-1375.

ANH

Monaco F, Lei C, Bonizzoni MA, Efremov S, Morselli F, Guarracino F, et al. A Randomized Trial of Acute Normovolemic Hemodilution in Cardiac Surgery. *N Engl J Med.* 2025;393(5):450-460.

Introduction

Acute normovolaemic haemodilution has an appealing logic. Whole blood is removed before cardiopulmonary bypass, kept outside the circuit and returned after bypass and heparin reversal. During the operation, bleeding occurs at a lower red-cell concentration, while the stored autologous blood preserves red cells, platelets and coagulation proteins for later use. If the patient tolerates the initial reduction in circulating red-cell mass and intravascular volume is managed appropriately, fewer donor products might be required.

The same manoeuvre creates potential disadvantages. Removing blood can provoke hypotension or anaemia, particularly in patients with limited cardiac reserve. Replacing the volume with crystalloid adds further dilution. If the blood has to be returned before bypass ends, part of the intended protection from circuit exposure is lost. A physiological rationale therefore cannot establish the balance of clinical benefit and harm.

Earlier small trials and pooled analyses suggested reduced transfusion, and ANH became one option within cardiac surgical blood-management programmes. The evidence was heterogeneous, including variation in withdrawal volume, fluid replacement, transfusion thresholds and background practice. The multinational ANH trial provided a much larger contemporary test of a reproducible strategy: withdraw at least 650 mL before heparinisation and return it after bypass, compared with usual care without ANH.¹

Synopsis

The ANH trial was a pragmatic, individually randomised study at thirty-two centres in eleven countries across Europe, Asia and the Americas. Recruitment took place from April 2019 to December 2024. The Italian Ministry of Health funded the trial and had no role in its design, data collection, analysis or writing. An independent monitoring committee reviewed interim data. All patients provided written consent before enrolment.¹

Adults scheduled for elective cardiac surgery with cardiopulmonary bypass underwent initial screening. Major exclusions included unstable coronary disease, emergency surgery, a critical preoperative state and inadequate interruption of anticoagulant or antiplatelet treatment; low-dose aspirin was permitted. A second assessment after anaesthetic induction excluded patients in whom withdrawing the target volume was judged likely to cause instability or excessive anaemia.

That second assessment incorporated predicted haematocrit after withdrawal and subsequent circuit priming. The supplementary eligibility criteria specified concern about a pre-bypass haematocrit below 30% or an on-bypass value below 24%. These were selection safeguards, not the same as the lower suggested transfusion triggers during the operation. Clinicians could also exclude patients on broader medical grounds. The study consequently selected patients considered able to tolerate haemodilution.²

A web-based system assigned participants equally using computer-generated permuted blocks stratified by centre. The attending anaesthetist and operating-room staff performing the intervention knew the allocation. Patients, postoperative intensive care and ward staff, outcome assessors, follow-up personnel and statisticians were intended to remain masked. Treatment information was accessible for safety, although the supplementary report states that unmasking was not required.

In the intervention group, at least 650 mL of whole blood was withdrawn before heparin administration, usually through a large-bore central catheter. Standard autologous collection bags containing anticoagulant were used, and blood was stored at room temperature. At the anaesthetist's discretion, up to 3 mL of crystalloid could be given for each millilitre removed. The collected blood was returned after bypass weaning and protamine reversal. If anaemia developed earlier, clinicians were instructed to use the stored blood before donor red cells.

Control patients received usual local care without ANH. Suggested haematocrit transfusion thresholds were below 28% before bypass, below 20% during bypass, below 25% immediately after weaning and below 27% during the postoperative admission. These thresholds were recommended rather than mandatory. Other conservation measures varied: twenty-two centres routinely used prophylactic tranexamic acid, twenty-one used autologous circuit priming, and twenty-five used viscoelastic-based haemostasis algorithms.^{1,2}

The primary outcome was receipt of at least one allogeneic red-cell unit from randomisation to hospital discharge. Secondary outcomes included death within thirty days after surgery or during the surgical admission, bleeding complications, ischaemic

events and acute kidney injury. Safety measures included cardiogenic shock, prolonged vasoactive support, mechanical circulatory support, sepsis and postoperative haematocrit. Follow-up was undertaken by masked personnel using telephone contact and available clinical records.

The investigators assumed a 35% control transfusion rate and sought a 20% relative reduction. Approximately 2,000 participants provided 90% power with a two-sided alpha of 0.05, allowing for interim analyses, withdrawal and rounding. Three planned reviews used stringent early efficacy boundaries. The main comparison followed allocation and used available observations without initial imputation; missing-data sensitivity analyses, a per-protocol comparison and a post hoc centre-adjusted analysis were also presented. Secondary and subgroup intervals were not adjusted for multiple comparisons.

Of 7,205 screened patients, 2,358 consented and 2,010 were randomised: 1,010 to ANH and 1,000 to usual care. Primary-outcome data were available for 1,005 and 997, respectively. Median ages were 59 and 61 years, approximately one-fifth were women, and baseline haemoglobin was generally high, with medians of 14 and 15 g/dL. About 61% underwent valve surgery, 39% coronary bypass and 10% ascending aortic or arch surgery; these categories could overlap.¹

Median withdrawal volume was 650 mL, with an interquartile range of 650–700 mL. Forty-five patients with available data had less than 650 mL removed. Stored blood was returned before bypass in seven patients and during bypass in sixty-nine. Two collections could not be reinfused because of bag rupture or coagulation. Eighteen patients underwent plasmapheresis rather than whole-blood withdrawal. Crossovers occurred in thirty-five ANH patients and thirty-two controls, approximately 3.3% overall.

Median pre-bypass crystalloid administration was 1,000 versus 700 mL. Median bypass duration was 102 versus 98 minutes. Recorded intraoperative hypotension occurred in 114/1,005 ANH patients, 11.3%, and 46/994 controls, 4.6%. Haematocrit values after heparin and after protamine were lower with ANH, while the lowest postoperative intensive care haematocrit was similar. The intervention therefore produced measurable haemodilution and additional circulatory disturbance despite the screening safeguards.^{1,2}

At least one donor red-cell transfusion was given to 274/1,005 ANH patients, 27.3%, and 291/997 controls, 29.2%. The relative risk was 0.93 (95% CI 0.81–1.07; P=0.34). The observed absolute difference was approximately two percentage points. Per-protocol, missing-data and centre-adjusted analyses did not materially change the interpretation.

Among patients who received red cells, the median number of units was two in both groups, with interquartile ranges of 1–4 and 1–3. Total donor red-cell use was 880 versus 791 units, so the small numerical reduction in the proportion exposed did not translate into lower total consumption. Intraoperative exposure alone was less frequent with ANH, 5.0% versus 7.1%, but this exploratory difference was not sustained as a clear reduction across the hospital admission.²

Reoperation for bleeding occurred in 38/1,004 versus 26/995 patients, 3.8% versus 2.6%, relative risk 1.45 (95% CI 0.89–2.37). Median twelve-hour drain output was 290 versus 300 mL. Acute kidney injury occurred in 8.5% versus 8.9%, relative risk 0.95 (95% CI 0.71–1.26). Death within thirty days or during the surgical admission occurred in fourteen versus sixteen patients, relative risk 0.87 (95% CI 0.42–1.76). There were no clear differences in major ischaemic events, prolonged vasoactive support, ventilation or length of stay.

Most subgroup estimates were consistent with the overall result. In patients whose bypass lasted no more than 102 minutes, the relative risk of transfusion was 0.72 (95% CI 0.56–0.94), whereas the estimate for longer bypass was 1.04 (95% CI 0.88–1.22). These exploratory findings were presented without multiplicity-adjusted intervals and did not establish a reliable treatment rule.²

Critique

The principal strength is the scale and geographical breadth of the randomised comparison. This one trial enrolled almost as many participants as the twenty-nine trials in an influential earlier meta-analysis. Central randomisation, very little missing primary-outcome information and broadly consistent sensitivity analyses make the neutral finding difficult to dismiss as a simple failure of trial execution.

The result is not proof that the strategies are equivalent. The primary confidence interval remains compatible with a relative reduction in transfusion approaching 19%, as well as a small increase. The observed control-event rate was below the 35% anticipated during planning, reducing the information available for detecting a given relative effect. Nevertheless, the study did not demonstrate its intended reduction, and the point estimate was considerably smaller than the benefit suggested by earlier pooled evidence.

Clinical interpretation should use the entire transfusion pattern. A binary outcome treats one donor unit and many donor units as equivalent exposure. That could miss a useful reduction in quantity among patients who still need some blood. Here, however, the

secondary data do not provide a convincing quantity-saving signal: transfused patients received a median of two units in both groups, and the ANH group used more units overall. Those totals are descriptive and may be influenced by a small number of heavy users, but they do not support an alternative claim that a major conservation benefit was hidden by the primary endpoint.

The medians require their own denominator warning. They describe patients who were transfused, not all randomised patients. Comparing only recipients conditions on an outcome that treatment could itself influence. Such summaries help characterise exposure but should not be interpreted as a clean randomised comparison of units per patient. Whole-cohort transfusion incidence and total resource use provide complementary information, neither showing the expected broad benefit.

The timing results suggest that ANH may alter when donor blood is given more readily than whether it is given during the admission. Fewer intervention patients received intraoperative donor red cells, yet later exposure reduced that difference. This is an exploratory observation across several time windows, not a separately confirmed benefit. It also illustrates why stopping follow-up at the operating-room door could have produced a more favourable but less complete assessment of conservation.

The study population was selected for tolerance rather than maximum transfusion risk. Unstable coronary disease, emergency procedures, ongoing instability and insufficient haematological reserve excluded many candidates. Baseline haemoglobin was high and women were underrepresented, partly reflecting the practical difficulty of safely removing a fixed volume in patients with smaller blood volumes or lower haemoglobin. The result applies most directly to stable elective patients meeting these safeguards.

It would be incorrect, however, to describe all complex surgery as excluded. Valve, aortic and combined procedures were represented. In subsequent correspondence, the investigators noted that 647 patients had bypass lasting more than two hours and clarified the substantial proportion undergoing operations beyond isolated coronary or single-valve surgery. Higher-risk and complex patients were not interchangeable with the haemodynamically unstable patients excluded for safety.^{3,4}

Withdrawal volume is the most plausible unresolved intervention question. Most patients received very close to the minimum 650 mL, rather than a wide range of weight-based volumes. For a large adult, this removes a relatively modest fraction of circulating blood. A letter argued that larger-volume haemodilution might be more effective. The reply explained that 650 mL was chosen to balance earlier efficacy signals

against hypotension and anaemia. Neither argument establishes the clinical effect of a larger-volume strategy in this population.^{4,5}

Dose escalation cannot be assumed to preserve safety. Hypotension was already more frequent in the ANH group, and some stored blood had to be returned before the intended time. These findings are relevant even though major postoperative safety outcomes did not differ clearly. A strategy that removes more blood might protect more cells from bypass exposure, but it could also increase circulatory stress or force early reinfusion. A direct comparison is needed to determine that balance.

The label 'normovolaemic' also should not imply that intravascular volume was experimentally held constant. Crystalloid replacement was discretionary, up to a specified maximum, and other haemodynamic manoeuvres followed local practice. Fluid distribution, vasodilatation and the timing of replacement complicate the relationship between the volume collected and the volume remaining within the circulation. The trial tested a pragmatic withdrawal-and-replacement policy, not a tightly controlled physiological state.

Blinding was more extensive than in many procedural trials, but the clinician performing ANH necessarily knew allocation. That anaesthetist also made early haemodynamic and transfusion decisions. Masked postoperative care and outcome collection reduce downstream assessment bias, while leaving the possibility of allocation-influenced intraoperative decisions. Because the primary outcome records actual donor transfusion rather than a judgement score, ascertainment is relatively robust, but the act being recorded remains a clinical decision.

Allowing local transfusion policies improves relevance to ordinary practice and introduces heterogeneity. A fixed haemoglobin or haematocrit threshold can dominate whether a small difference in red-cell mass becomes a transfusion. Background use of tranexamic acid, circuit priming and coagulation testing varied substantially. Centre-stratified randomisation should balance these practices between groups within each hospital, so variation alone does not invalidate the result. It does limit claims about the effect within a fully standardised conservation programme.

The centre-adjusted sensitivity analysis was consistent with the primary result, and crossover was modest in both directions. Nonadherence could dilute a true effect, but the per-protocol analysis did not reveal a convincing benefit. Removing patients who could not tolerate withdrawal would also create a selected comparison and could conceal an important practical limitation. The ability to deliver a procedure safely is part of its effectiveness, not merely an inconvenience to be excluded from interpretation.

The short-bypass subgroup is particularly vulnerable to overinterpretation. Actual bypass duration is known only after randomisation and reflects the operation and its complications. It cannot serve as a simple preoperative selection rule. A favourable confidence interval in one of several subgroups is also not, by itself, proof that the treatment effects differ. The reply's suggestion to consider ANH in shorter procedures should be regarded as a hypothesis rather than a demonstrated indication.⁴

Safety language needs similar discipline. Death, stroke, myocardial infarction and dialysis were uncommon, producing wide intervals. The numerical increase in reoperation is not proof of harm, but its interval also does not exclude a clinically important increase. The main report's general reassurance about safety should be read alongside the clear descriptive excess in intraoperative hypotension. An absence of established major harm does not create a reason to routinely perform a procedure without established benefit.

The physiological and resource rationale remains relevant to future research. Preserving platelet-rich plasma while promptly returning red cells, using a weight-based withdrawal, or targeting a particular surgical phenotype could be different interventions. The small plasmapheresis subgroup in this trial cannot validate any of them. Future studies would need explicit delivery targets, appropriate safety monitoring and outcomes that capture both donor exposure and total treatment burden.

For present interpretation, ANH is a strong corrective to confidence built from many small studies. Its findings do not prove that every possible version of autologous collection is ineffective. They do remove support for assuming that routine withdrawal of approximately 650 mL in eligible elective cardiac surgical patients meaningfully reduces hospital-wide donor red-cell use. The distinction preserves a research question without treating an untested variation as an established rescue for the tested strategy.

Body of Evidence

The 2017 Barile meta-analysis combined twenty-nine randomised trials with 2,439 participants. It estimated approximately 0.79 fewer donor red-cell units with ANH and a lower proportion receiving allogeneic transfusion, pooled risk ratio 0.74 (95% CI 0.62–0.87). Heterogeneity was substantial: approximately 95% for unit use and 73% for transfusion incidence. These results explain the earlier enthusiasm, but also show why the pooled average could be unstable across techniques and settings. The new trial is an independent, much larger test within one defined contemporary strategy.⁶

Ming and colleagues directly compared larger- and moderate-volume ANH in 110 patients at one university hospital. Mean collections were approximately 886 and 395

mL, with transfusion in 23.6% versus 41.8%. This supports the plausibility of a volume effect, but both groups underwent ANH. It does not establish that larger-volume collection is superior to no collection, and its small size limits assessment of uncommon harms. It should inform a new trial rather than override the larger usual-care comparison.⁷

TRICS III provides a more secure example of reducing donor exposure through a different mechanism. In 5,243 higher-risk cardiac surgical patients, a restrictive haemoglobin threshold of 7.5 g/dL was noninferior to a more liberal strategy for death, myocardial infarction, stroke or new dialysis. Red-cell exposure was 52.3% versus 72.6%. The population and intervention differ from ANH, but the study demonstrates that transfusion decisions themselves are a major determinant of product use and that conservation can be assessed against clinical safety outcomes.⁸

ATACAS evaluated tranexamic acid in coronary surgery. Among 4,631 operated patients with outcome data, it reduced transfused blood products and reoperation for major bleeding or tamponade, 1.4% versus 2.8%. It did not clearly change the composite of death and thrombotic complications, and seizures increased from 0.1% to 0.7%. This is another reminder that blood-conservation efficacy and overall safety are separate questions. It also provides context for the varying background antifibrinolytic practice in ANH.⁹

PRICE-2 tested blood withdrawal in a different physiological setting. Among 446 analysed liver-resection patients, hypovolaemic phlebotomy without volume replacement reduced thirty-day donor red-cell exposure from 16% to 8%, adjusted risk ratio 0.47 (95% CI 0.27–0.82). Lower venous pressure during hepatic transection is a different proposed mechanism from protecting blood during bypass. Its positive result cannot be transferred to cardiac surgery, where deliberate hypovolaemia may have a different risk–benefit balance.¹⁰

The combined evidence supports evaluating each conservation method on its own delivery conditions and clinical outcomes. ANH challenges a widely plausible practice, while TRICS III and ATACAS show that some other approaches can reduce donor use with defined benefits and risks. The appropriate response is to retain evidence-based blood management and reconsider the routine contribution of this particular haemodilution strategy, rather than assuming that all conservation measures are interchangeable.

Should we routinely perform acute normovolaemic haemodilution before cardiac surgery?

The tested strategy does not support routine use: withdrawing approximately 650 mL did not clearly reduce hospital red-cell transfusion. Larger-volume approaches and the short-bypass subgroup remain unconfirmed, with haemodynamic tolerance an important constraint.

References

1. Monaco F, Lei C, Bonizzoni MA, Efremov S, Morselli F, Guarracino F, et al. A Randomized Trial of Acute Normovolemic Hemodilution in Cardiac Surgery. *N Engl J Med*. 2025;393(5):450-460.
2. Monaco F, Lei C, Bonizzoni MA, Efremov S, Morselli F, Guarracino F, et al. Supplementary methods and results to A Randomized Trial of Acute Normovolemic Hemodilution in Cardiac Surgery. *N Engl J Med*. 2025;393(5):450-460.
3. Tamura T. Acute Normovolemic Hemodilution in Cardiac Surgery. *N Engl J Med*. 2025;393(19):1963.
4. Monaco F, Bonizzoni MA, Landoni G. Acute Normovolemic Hemodilution in Cardiac Surgery. Reply. *N Engl J Med*. 2025;393(19):1963-1964.
5. Ahmed AA. Acute Normovolemic Hemodilution in Cardiac Surgery. *N Engl J Med*. 2025;393(19):1963.
6. Barile L, Fominskiy E, Di Tomasso N, Alpizar Castro LE, Landoni G, De Luca M, et al. Acute Normovolemic Hemodilution Reduces Allogeneic Red Blood Cell Transfusion in Cardiac Surgery: A Systematic Review and Meta-analysis of Randomized Trials. *Anesth Analg*. 2017;124(3):743-752.
7. Ming Y, Zhang F, Yao Y, Cheng Z, Yu L, Sun D, et al. Large volume acute normovolemic hemodilution in patients undergoing cardiac surgery with intermediate-high risk of transfusion: A randomized controlled trial. *J Clin Anesth*. 2023;87:111082.
8. Mazer CD, Whitlock RP, Fergusson DA, Hall J, Belley-Cote E, Connolly K, et al. Restrictive or Liberal Red-Cell Transfusion for Cardiac Surgery. *N Engl J Med*. 2017;377(22):2133-2144.
9. Myles PS, Smith JA, Forbes A, Silbert B, Jayarajah M, Painter T, et al. Tranexamic Acid in Patients Undergoing Coronary-Artery Surgery. *N Engl J Med*. 2017;376(2):136-148.
10. Martel G, Carrier FM, Wherrett C, Lenet T, Mallette K, Brousseau K, et al. Hypovolaemic phlebotomy in patients undergoing hepatic resection at higher risk of blood loss (PRICE-2): a randomised controlled trial. *Lancet Gastroenterol Hepatol*. 2025;10(2):114-124.

Kougias P, Sharath SE, Zhan M, Carson JL, Norman LE, Mi Z, et al. Liberal or Restrictive Postoperative Transfusion in Patients at High Cardiac Risk: The TOP Randomized Clinical Trial. JAMA. 2025;334(24):2197-2207.

Introduction

Postoperative anaemia creates competing concerns in patients with cardiovascular disease. A lower haemoglobin concentration reduces oxygen-carrying capacity at a time when pain, inflammation and surgical stress may increase myocardial demand. Transfusion can improve oxygen delivery, but also introduces volume, donor exposure and other potential complications. The optimum decision depends on more than whether either intervention has a plausible physiological benefit.

Evidence supporting restrictive transfusion is substantial, but it comes from different populations and different thresholds. A stable patient with peripheral arterial disease after vascular surgery is not identical to someone with an acute myocardial infarction, nor to a patient whose coronary circulation has just been surgically revascularised. Applying a single conclusion across all three settings risks overlooking clinically important differences.

The Transfusion Trigger after Operations in High Cardiac Risk Patients trial, TOP, addressed postoperative anaemia after major noncardiac surgery. It compared a liberal haemoglobin trigger of 10 g/dL with a restrictive trigger of 7 g/dL in patients selected for high baseline cardiovascular risk. The question was whether the additional transfusion prevented death or major ischaemic events, rather than simply whether it raised haemoglobin.¹

Synopsis

TOP was a parallel-group, individually randomised superiority trial at sixteen United States Veterans Affairs medical centres. Recruitment ran from February 2018 to March 2023. The Veterans Affairs Office of Research and Development funded and coordinated the work, with independent data-monitoring and endpoint committees. Participants provided consent and were intended to remain masked to allocation; treating investigators were unmasked.¹

Eligible adults had undergone a qualifying major vascular or general surgical operation and had a history indicating cardiovascular risk, such as ischaemic heart disease, myocardial infarction, peripheral arterial disease, stroke or transient ischaemic attack. Peripheral arterial disease procedures included bypasses, endarterectomy and major lower-limb amputation. The qualifying general operations were broad, but ultimately constituted a small proportion of the cohort.

Patients were randomised when an inpatient postoperative haemoglobin measurement fell below 10 g/dL within fifteen days of surgery. Exclusions included inability or unwillingness to receive transfusion, relevant hereditary anaemia or bleeding disorders, previous adverse blood-product reactions and persistent haemodynamic instability or cardiogenic shock under the specified criteria. Preoperative and intraoperative transfusion were outside the assigned strategy. This was a postoperative threshold trial, not a comparison of resuscitation during uncontrolled surgical haemorrhage.^{1,2}

A central telephone system used block randomisation stratified by site and Revised Cardiac Risk Index class. In the liberal group, the initial prescription was one red-cell unit for haemoglobin at least 8.5 g/dL but below the treatment threshold, two units for 7.5–8.4 g/dL, and three units below 7.5 g/dL. The restrictive group received one unit for haemoglobin 5.5–6.9 g/dL and two units below 5.5 g/dL. Subsequent treatment maintained haemoglobin at or above 10 or 7 g/dL, respectively, until discharge or thirty days after randomisation.

The assigned algorithm could be paused for an adverse transfusion reaction, a postrandomisation myocardial infarction or rapid bleeding. Haemoglobin and other laboratory measurements were scheduled during the early postoperative period, with electrocardiograms at defined time points. Staff reviewed measurements below the assigned threshold for appropriate transfusion, spontaneous recovery, a permitted exception or a deviation. Follow-up occurred at thirty and ninety days, with mortality additionally assessed at one year.

The primary outcome was all-cause death or a major ischaemic event within ninety days: myocardial infarction, coronary revascularisation, acute kidney failure or ischaemic stroke. Two masked committee members adjudicated each primary event. Acute kidney failure required a severe functional deterioration under the specified RIFLE criteria; it did not include every small postoperative creatinine increase.

Five secondary outcomes included infectious complications, cardiac complications other than myocardial infarction, the thirty-day primary composite, one-year mortality and postrandomisation hospital stay. The non-infarction cardiac composite comprised new

arrhythmia requiring treatment, new or worsening heart failure and nonfatal cardiac arrest. These secondary events were not independently adjudicated. A further composite adding myocardial infarction to those cardiac complications was post hoc.^{1,2}

The planned sample of 1,520 allowed 90% power to detect a 25% relative reduction in the primary outcome, assuming event rates of 30% with restriction and 22.5% with liberal transfusion and 5% attrition. Recruitment stopped at 1,428 because further funding was unavailable. Interim reviews were planned at one-quarter, one-half and three-quarters of completed primary follow-up. The principal analysis compared available outcomes by assigned group. Secondary and individual-component analyses used a more stringent alpha of 0.01 and 99% confidence intervals, while the primary comparison used 95% intervals.

Of 44,183 veterans assessed, 3,022 consented and 1,428 were randomised, 714 per group. Four were subsequently excluded because of duplicate trial participation or non-veteran status, leaving 712 per group in the analysis cohort. Mean age was approximately 70 years and 97.8% were male. Peripheral arterial disease affected 72%, coronary disease 59%, diabetes 49% and heart failure 24%. Vascular operations accounted for 91.1% of procedures.¹

The strategies produced a clear treatment contrast. Median postrandomisation red-cell use was two units with liberal treatment and zero with restriction. In the liberal group, 666/712 patients received blood; in the restrictive group, 164/712 did. Reported postrandomisation totals, excluding transfusion during subsequent operations, were 1,878 versus 307 units. Mean haemoglobin on day five was 10.8 versus 8.8 g/dL, a difference of 2.0 g/dL. Major deviations were reported in approximately 9% versus 4%.

Primary-outcome information was available for 670 liberal-group and 700 restrictive-group patients. The composite occurred in 61/670, 9.1%, versus 71/700, 10.1%, relative risk 0.90 (95% CI 0.65–1.24). Adjustment for age, cardiac risk and the postoperative haemoglobin nadir gave a similar estimate, as did time-to-event analysis. Multiple imputation under missing-at-random and several missing-not-at-random scenarios did not materially change the result.^{1,2}

Ninety-day mortality was 4.6% versus 4.7%. Myocardial infarction occurred in 3.4% versus 4.1%, coronary revascularisation in 1.2% versus 1.9%, severe acute kidney failure in 1.7% versus 2.1%, and stroke in 0.3% versus 0.4%, using the available denominator for each component. None established a clear difference. One-year mortality was 14.1% versus 15.0%, relative risk 0.94 (99% CI 0.67–1.32). Median postrandomisation hospital stay was five days in both groups.

The prespecified non-infarction cardiac composite occurred in 38/647 liberal-group patients, 5.9%, and 67/678 restrictive-group patients, 9.9%, relative risk 0.59 (99% CI 0.36–0.98). Heart failure and arrhythmia counts both favoured liberal transfusion, although neither component alone had a precise effect estimate. Adding myocardial infarction in the post hoc composite gave 8.2% versus 12.5%, relative risk 0.66 (99% CI 0.43–1.01). Infectious complications were similar.

Thirteen transfused liberal-group patients and two transfused restrictive-group patients had an adverse transfusion reaction, approximately 2.0% and 1.2% of recipients. Voluntary withdrawal was more frequent with liberal treatment, thirty-one versus five patients, commonly because patients preferred discharge to continuing transfusions required by the strategy. Primary-outcome data were incomplete for forty-two versus twelve patients overall.^{1,2}

Critique

TOP addresses a meaningful gap between general restrictive-transfusion evidence and concern about myocardial oxygen supply in postoperative patients with established vascular disease. Its randomised design, marked haemoglobin separation and independent adjudication of the primary components are substantial strengths. Lack of treatment separation is not a convincing explanation for the neutral primary comparison: the liberal strategy delivered much more blood and achieved a clearly higher haemoglobin concentration.

The main limitation is precision. The actual control-event rate was approximately 10%, compared with the 30% anticipated during planning. Recruitment also stopped modestly short of target, but the event-rate shortfall was the more substantial loss of information. Only 132 primary events were observed. The confidence interval permits a 35% relative reduction or a 24% increase with liberal transfusion. Those possibilities are not clinically interchangeable.

This was a superiority trial. Failure to establish superiority of liberal treatment does not demonstrate noninferiority of the restrictive strategy, because no noninferiority margin was tested. Nor does it establish equivalence. The result weakens the case for routinely giving enough blood to maintain 10 g/dL in every similar patient, while leaving meaningful uncertainty about the consequences of allowing haemoglobin to fall towards 7 g/dL in particular patients.

The authors discuss the very large sample that would be required to reliably detect the observed small point estimate. Such a calculation should not be used to decide that all

remaining differences are unimportant. The observed estimate is itself uncertain, and a future study could target a larger clinically important effect or a more informative population. The practical interpretation rests on the confidence interval and the treatment burden, not on the impossibility of confirming one assumed exact effect size.

The primary composite is broad. Death, infarction, revascularisation, stroke and severe renal failure differ in frequency, mechanism and responsiveness to transfusion. A benefit concentrated in myocardial stress could be diluted by unrelated deaths or renal events. Conversely, a composite offers more statistical information than any rare component alone. The generally favourable direction of several point estimates is compatible with benefit, but it is not independent replication: the outcomes arise in the same patients and can overlap.

The secondary cardiac result merits attention without displacing the primary endpoint. It was prespecified, clinically coherent and met the trial's more stringent 0.01 threshold, with a 99% interval just excluding no effect. It therefore deserves more consideration than an isolated unplanned P value. However, it was one of several secondary outcomes, had fewer complete observations, and lacked independent masked adjudication. Its apparent strength should be weighed against those differences in ascertainment.

Heart failure and treatment-requiring arrhythmia can be influenced by clinical recognition and management. Investigators knew the transfusion strategy, and the criteria incorporated initiation or escalation of treatment. Although there were defined diagnostic requirements, knowledge of anaemia and allocation could affect investigation or attribution. This concern is less pronounced for the separately adjudicated primary events. The secondary finding is a reason for further investigation, not conclusive evidence that liberal transfusion prevents postoperative heart failure.

Adding infarction to the cardiac composite produced an estimate in the same direction but a 99% interval crossing no effect. That does not mean infarction reversed a proven benefit, nor that the original composite was invalid. It reflects a changed endpoint, event distribution and uncertainty. The expanded analysis was post hoc and should retain that status. Its favourable point estimate cannot be promoted into a new primary result.

The apparently paradoxical heart-failure finding also illustrates the limitations of mechanistic shortcuts. Additional transfused volume could worsen congestion, while persistent anaemia could increase cardiac work and impair oxygen supply. Both mechanisms are plausible. Fluid administration and diuretic management were not systematically characterised sufficiently to establish which pathway explained the

observed events. The trial does not show that volume is irrelevant or that raising haemoglobin necessarily improves ventricular function.

Missing data were limited overall but unevenly distributed. Forty-two liberal-group patients and twelve restrictive-group patients lacked complete primary-outcome information. More liberal-group patients withdrew because the transfusion strategy conflicted with their desire to leave hospital. This is potentially related to treatment and postoperative circumstances rather than random loss. The imbalance matters when the observed absolute primary difference is only about one percentage point.

The imputation analyses are reassuring within their assumptions. They included twenty imputed datasets and sensitivity scenarios altering the odds of events among missing observations. Similar estimates under these models reduce concern that a modest departure from the assumed missingness mechanism explains the finding. They cannot prove the absence of bias under every plausible pattern, particularly if unobserved outcomes differed between arms in a way not captured by the models. The chapter therefore retains the actual denominators rather than describing outcome data as complete.

The intervention was also more than a haemoglobin number. Liberal treatment often involved initial blocks of two or three units, followed by a requirement to maintain the target until discharge. This generated substantially greater exposure and may have affected acceptability. It differs from a flexible strategy in which a clinician gives one unit, reassesses symptoms and stops at an intermediate concentration. TOP cannot identify the best threshold between 7 and 10 g/dL, because it did not randomise those alternatives.

The mean haemoglobin separation on day five describes patients still measured at that point. It is a useful sign of successful delivery, but should not be treated as the identical contrast experienced by every participant for thirty days. Median postrandomisation stay was five days, and the intervention ended at discharge. A ninety-day endpoint therefore assesses consequences of a relatively short inpatient policy, with subsequent anaemia and transfusion care outside the assigned intervention.

Applicability is strongest for older men undergoing vascular surgery in an integrated Veterans Affairs system. Only thirty-one women were analysed, and general surgery accounted for fewer than one in ten procedures. Major oncological resections and other operations were poorly represented. The large screening denominator also reflects specific eligible procedures and the requirement for postoperative anaemia, rather than a sample of all surgical patients with cardiovascular risk.

High baseline cardiac risk should not be confused with an acute coronary syndrome. Many participants qualified through peripheral arterial or cerebrovascular disease, and cardiac risk classes were heterogeneous. A postrandomisation myocardial infarction allowed suspension of the assigned algorithm. TOP therefore cannot establish how to transfuse a patient who is already experiencing acute myocardial ischaemia. That question requires the distinct acute-infarction evidence.

The transfusion-reaction comparison also needs a denominator distinction. The reported percentages were among patients who received blood, a group selected partly by randomised strategy. Far more liberal-group patients were exposed, so a similar reaction rate per recipient would still imply a different burden across all assigned patients. Conversely, a small number of recognised reactions does not capture every possible effect of transfusion or anaemia. Patient outcomes and resource use must be considered together.

The accompanying editorial advocates integrating symptoms, haemodynamic state and postoperative trajectory with haemoglobin. That is a reasonable interpretation of the unresolved heterogeneity, but TOP did not test a biomarker-guided or symptom-adaptive algorithm. Proposed refinements should be presented as clinical considerations and research questions, rather than treatments shown superior by this trial.³

The practical contribution is therefore balanced. Routine maintenance at 10 g/dL consumed considerably more blood without a demonstrated improvement in the primary ninety-day endpoint. At the same time, the trial's precision and secondary cardiac signal do not support declaring the lower threshold uniformly safe across every cardiovascular phenotype. It advances the postoperative evidence while preserving the need to distinguish stable chronic disease from active ischaemia and to assess the patient's whole clinical course.

Body of Evidence

FOCUS enrolled 2,016 patients after hip-fracture surgery with cardiovascular disease or risk factors and haemoglobin below 10 g/dL. Liberal transfusion at 10 g/dL did not improve death or inability to walk independently at sixty days compared with transfusion for symptoms or haemoglobin below 8 g/dL: 35.2% versus 34.7%, odds ratio 1.01 (95% CI 0.84–1.22). This is relevant postoperative evidence, but its restrictive strategy included symptoms and a higher numerical threshold than TOP. It cannot by itself validate waiting until 7 g/dL.⁴

TRICS III studied 5,243 patients at moderate or high risk undergoing cardiac surgery. A restrictive 7.5 g/dL threshold was noninferior to more liberal thresholds for death, infarction, stroke or new dialysis, while reducing transfusion exposure. The primary event rates were 11.4% versus 12.5%. This provides a formal noninferiority comparison, unlike TOP, but involves cardiac surgery and a different relationship between the operation and coronary blood supply.⁵

REALITY addressed acute myocardial infarction with anaemia. Among 668 randomised patients, a restrictive threshold of 8 g/dL met the trial's noninferiority criterion against 10 g/dL for thirty-day major cardiovascular events. Events occurred in approximately 11% versus 14%, with a relative risk of 0.79 and an upper one-sided 97.5% confidence limit of 1.19. Even that result allowed clinically important harm within the accepted margin and did not settle the acute-ischaemia question.⁶

MINT supplied a larger acute-infarction comparison, with 3,504 patients in its primary analysis. Death or recurrent infarction occurred in 16.9% with a restrictive 7–8 g/dL strategy and 14.5% with liberal treatment at 10 g/dL, relative risk 1.15 (95% CI 0.99–1.34; $P=0.07$). The result did not establish liberal superiority but raised concern about harm from restriction. The relative-risk direction here is restrictive versus liberal, the reverse of TOP's reported primary estimate.⁷

An individual-patient meta-analysis of four acute-infarction trials, including 4,311 patients, estimated a relative risk of 1.13 (95% CI 0.97–1.30) for thirty-day death or infarction with restriction. Cardiac death favoured liberal treatment, while the primary composite remained uncertain. This pooled analysis incorporates MINT and REALITY rather than providing an independent additional cohort. It reinforces the distinction between active infarction and stable postoperative cardiovascular risk.⁸

Taken together, these studies do not identify one universally appropriate trigger. They show how thresholds, rescue provisions, surgical context and acute ischaemia change the question. TOP adds direct evidence for major noncardiac surgery with established cardiovascular disease: much greater transfusion exposure at 10 g/dL did not produce a confirmed primary benefit, but the remaining uncertainty is too broad to equate that result with universal endorsement of 7 g/dL.

Should postoperative patients at high cardiac risk routinely be transfused to 10 g/dL?

TOP did not establish a benefit from routine maintenance at 10 g/dL despite substantially greater blood use. It also did not prove equivalence or settle transfusion decisions in active myocardial ischaemia.

References

1. Kougias P, Sharath SE, Zhan M, Carson JL, Norman LE, Mi Z, et al. Liberal or Restrictive Postoperative Transfusion in Patients at High Cardiac Risk: The TOP Randomized Clinical Trial. *JAMA*. 2025;334(24):2197-2207.
2. Kougias P, Sharath SE, Zhan M, Carson JL, Norman LE, Mi Z, et al. Supplementary methods and results to Liberal or Restrictive Postoperative Transfusion in Patients at High Cardiac Risk: The TOP Randomized Clinical Trial. *JAMA*. 2025;334(24):2197-2207.
3. Jacobs JW, Bloch EM. Postoperative Transfusion in Patients at High Cardiac Risk: Evidence, Uncertainty, and Nuance. *JAMA*. 2025;334(24):2171-2173.
4. Carson JL, Terrin ML, Noveck H, Sanders DW, Chaitman BR, Rhoads GG, et al. Liberal or restrictive transfusion in high-risk patients after hip surgery. *N Engl J Med*. 2011;365(26):2453-2462.
5. Mazer CD, Whitlock RP, Fergusson DA, Hall J, Belley-Cote E, Connolly K, et al. Restrictive or Liberal Red-Cell Transfusion for Cardiac Surgery. *N Engl J Med*. 2017;377(22):2133-2144.
6. Ducrocq G, Gonzalez-Juanatey JR, Puymirat E, Lemesle G, Cachanado M, Durand-Zaleski I, et al. Effect of a Restrictive vs Liberal Blood Transfusion Strategy on Major Cardiovascular Events Among Patients With Acute Myocardial Infarction and Anemia: The REALITY Randomized Clinical Trial. *JAMA*. 2021;325(6):552-560.
7. Carson JL, Brooks MM, Hébert PC, Goodman SG, Bertolet M, Glynn SA, et al. Restrictive or Liberal Transfusion Strategy in Myocardial Infarction and Anemia. *N Engl J Med*. 2023;389(26):2446-2456.
8. Carson JL, Fergusson DA, Noveck H, Mallick R, Simon T, Rao SV, et al. Restrictive versus Liberal Transfusion in Myocardial Infarction - A Patient-Level Meta-Analysis. *NEJM Evid*. 2025;4(2):EVIDoa2400223.

CLIP-II

**Reade MC, Marks DC, Howe BD, Bailey MJ, Bannon PG, Eastwood GM, et al.
Cryopreserved vs Liquid-Stored Platelets for the Treatment of Surgical Bleeding:
The CLIP-II Randomized Noninferiority Clinical Trial. JAMA. 2026;335(7):600-608.
First published online 8 December 2025.**

Introduction

Platelets are difficult to make available everywhere they might be needed. Conventional storage lasts only five to seven days, creating a tension between maintaining adequate emergency stocks and discarding unused donations. A large cardiac centre can anticipate substantial demand; a small hospital or deployed military service may have too little routine use to support a reliable supply. A longer-lived product could therefore change access as well as reduce wastage.

Cryopreservation offers a potential solution. Platelets stored at -80°C with dimethyl sulfoxide can remain available for two years, then be thawed and resuspended in plasma. However, freezing changes the cells. Increased procoagulant surface activity may coexist with impaired aggregation, lower circulating increments and altered survival. A product can perform well in a coagulation assay without necessarily controlling clinical bleeding as effectively as conventional platelets.^{1,2}

CLIP-II tested whether the logistical advantages of cryopreserved platelets could be obtained without an unacceptable loss of haemostatic effectiveness. Cardiac surgery provided a setting with predictable platelet demand, measurable chest-drain bleeding and an established comparator. The trial was designed to establish noninferiority, rather than to require the new product to reduce bleeding beyond what conventional platelets achieved.¹

Synopsis

CLIP-II was an investigator-initiated, multicentre, double-blind, parallel-group randomised trial in eleven Australian tertiary hospitals. Recruitment ran from August 2021 to April 2024, with follow-up completed in July. Support came from the National Health and Medical Research Council, Australian Red Cross Lifeblood and the Australian Defence Force. Several investigators had roles within the organisations developing or supplying the product; the report describes their involvement rather than presenting the funders as entirely separate from trial conduct.

Adults undergoing cardiac surgery were eligible if the Australian Cardiac Surgery Platelet Transfusion score was at least one or their clinician expected platelet transfusion to be needed. Consent and randomisation occurred before surgery. Important exclusions included recent deep-vein thrombosis or pulmonary embolism, more than one lifetime episode, significant specified coagulopathy, and relevant anticoagulant or antiplatelet exposure without sufficient washout. Aspirin and prophylactic anticoagulation were not blanket exclusions. Women aged 18–55 years who were RhD-negative or of unknown status were excluded because the cryopreserved donations were RhD-positive.^{1,3}

A centrally generated allocation schedule used variable blocks and stratification by site. Cryopreserved platelets came from group O apheresis donations, frozen in dimethyl sulfoxide and reconstituted in ABO-matched or group AB plasma. Conventional platelets were either apheresis-derived or pooled from four whole-blood donations, suspended in plasma and additive solution, with compatibility determined by usual hospital practice. Thus the comparison involved complete manufactured products, including their suspension media and donor sources, rather than storage temperature alone.

Opaque covers concealed the study bags from patients and treating clinicians. Two staff members outside the patient's clinical care checked identity and compatibility. Clinicians decided whether and when platelets were required. The first study unit had to start in theatre or within twenty-four hours of ICU admission. Up to three study units could be administered during the ICU admission, after which further platelet treatment used open-label conventional products. Other haemostatic treatment remained at clinical discretion.

The primary outcome was chest-drain blood loss over the first twenty-four hours after ICU admission. The noninferiority margin allowed up to 20% greater bleeding with cryopreserved platelets. A sample of 202 transfused patients was calculated to provide more than 80% power, using previous blood-loss estimates and allowing for a skewed distribution. Five secondary and forty-two tertiary outcomes examined other bleeding measures, transfusion, physiological effects and recovery. Safety assessment included thrombosis, respiratory complications, infection and mortality.¹

Blood-loss comparisons used log-transformed data, with effects expressed as ratios of geometric means. Adjustment incorporated site and EuroSCORE II. Death was treated as a competing event in analyses of extubation and discharge. Secondary and tertiary confidence intervals were not adjusted for the large number of comparisons. One primary-outcome measurement was missing, and multiple imputation was used in a sensitivity analysis. An independent monitoring committee reviewed safety at the halfway point.

Of 879 patients meeting the inclusion criteria, 412 consented and 388 were randomised: 196 to cryopreserved and 192 to conventional platelets. Only 202 subsequently required and received study platelets, 104 and 98 respectively, and formed the clinical analysis population. Mean age was 64 years, 75.7% were male and 30.2% underwent nonelective surgery. Mean bypass duration was 188 minutes, median EuroSCORE II was approximately 3.4, and more than 70% of participants came from three hospitals.¹

The first transfusion was intraoperative in 86.6%. Median time from ordering platelets to starting transfusion was similar, sixty versus fifty-seven minutes. Cryopreserved units had been frozen within fifty hours of donation, stored for no more than two years and transfused within four hours of resuspension. Conventional units averaged 5.5 days old; only 15.6% were obtained by apheresis.

Early use of open-label conventional platelets was markedly imbalanced. Twenty-six cryopreserved-group patients, 25.0%, received them after the first study unit but before completing three study units, compared with three conventional-group patients, 3.1%. Across the entire treatment period, 42.3% versus 12.2% received open-label platelets. Median total platelet use, including study and open-label units, was two versus one units. The median platelet nadir in the first ICU day was nevertheless lower with cryopreservation, 103 versus 140 × 10⁹/L.^{1,3}

Primary-outcome geometric means were 605 mL with cryopreserved platelets and 535 mL with conventional platelets, ratio 1.13 (95% CI 0.96–1.34). The site- and severity-adjusted ratio was 1.11 (95% CI 0.94–1.31). Because both upper confidence limits exceeded 1.20, noninferiority was not established. Median blood loss was 610 versus 523 mL, with an estimated median difference of 87 mL (95% CI –50 to 280).¹

Several additional measures favoured conventional platelets. Total postoperative chest-drain loss had geometric means of 1,273 versus 972 mL, ratio 1.31 (95% CI 1.07–1.60). The ratio for intraoperative blood loss was 1.42 (95% CI 1.12–1.80). A major cardiac-surgical bleeding composite occurred in 31.7% versus 19.4%, although adjustment moved the difference's confidence interval across zero.

Red cells were given at some point to 74.0% versus 59.2%, and during the first twenty-four ICU hours to 52.9% versus 38.8%. Median total red-cell exposure through ICU discharge or day seven was three versus two units across all participants. Plasma was transfused in 75.0% versus 50.0% and cryoprecipitate in 71.2% versus 54.1%. These were additional blood-component requirements, beyond the assigned platelet preparation itself.¹

Ninety-day deaths numbered thirteen of 104, 12.5%, versus five of 98, 5.1%; the survival hazard ratio was 2.56 (95% CI 0.93–7.14). Reoperation for bleeding occurred in nine versus seven patients. Thrombotic events were uncommon, and no acute coronary syndrome, deep-vein thrombosis or prespecified dimethyl-sulfoxide complication was reported. Median ventilation duration was 25.5 versus 23.6 hours, ICU stay 3.8 versus 3.0 days, and hospital stay from first platelet transfusion 10.9 versus 9.1 days. Competing-risk analyses also favoured earlier extubation and discharge with conventional platelets.^{1,3}

A sensitivity analysis excluded the twenty-nine patients given early open-label platelets, leaving seventy-eight versus ninety-five participants. The primary geometric-mean ratio was 1.07 (95% CI 0.89–1.28), again failing to establish noninferiority. A separate post hoc analysis added blood already present in chest drains at ICU admission to the primary time window; its adjusted ratio was 1.24 (95% CI 1.06–1.44). These analyses addressed different departures from the principal comparison and retained their respective limitations.³

Critique

The trial asks an appropriate clinical question about a potentially valuable supply innovation. Random allocation, prospective consent, masked product administration and a readily measurable outcome are important strengths. Conduct across eleven hospitals improves relevance beyond one surgical team, while comparable delivery times reduce concern that simple thawing delay explains the findings. The study also moved beyond laboratory surrogates to bleeding, transfusion and recovery.

The primary interpretation requires care. A confidence interval crossing both one and the noninferiority margin is inconclusive for two different propositions: it neither establishes conventional statistical superiority of one product on the primary endpoint nor excludes an unacceptable loss of effectiveness with cryopreservation. Describing this as similar bleeding would discard the purpose of the design. Describing the primary endpoint alone as proof of inferiority would also go beyond the result.

The interval allows anything from a small reduction in bleeding to approximately one-third more bleeding with cryopreserved platelets. The latter possibility exceeds the chosen acceptable loss. Routine substitution therefore lacks the reassurance that the trial was intended to provide. This conclusion does not depend on treating every secondary finding as independently definitive.

The margin itself is a clinical judgement. Twenty per cent more chest-drain loss was considered acceptable in exchange for easier availability, but that trade-off differs between a tertiary hospital with conventional stocks and a remote facility with no platelets. A relative margin also translates into different absolute volumes as baseline bleeding changes. Blood loss was lower than anticipated from the pilot, while variability remained substantial, reducing the study's ability to narrow the uncertainty. Failure to meet the margin cannot be repaired by assuming that a larger study would necessarily favour noninferiority.

More fundamentally, conventional platelet transfusion has not been established against no platelets in an equivalent cardiac-surgical trial. The margin therefore does not represent a securely quantified fraction of an historical treatment effect. It is a pragmatic tolerance for additional bleeding. This limits what a successful noninferiority result would have proved and reinforces the need to examine clinically consequential secondary outcomes alongside drain volume.¹

The analysis population also deserves an explicit description. Almost half the randomised patients never received platelets and were excluded from the treatment comparison. Randomising in advance made blinded emergency delivery feasible. Because treatment was still concealed when clinicians decided that a first transfusion was necessary, selection was less vulnerable to knowledge of allocation than an unblinded postrandomisation decision would have been. Nevertheless, the reported result concerns patients actually transfused, not the policy effect of assigning all 388 surgical patients to one inventory strategy.

The initial masking procedures were credible, but masking may not have remained functionally intact. A lower platelet increment or persistent bleeding could have led clinicians to suspect the allocated product. The large imbalance in early open-label transfusion could represent recognition of reduced haemostatic effect, variation in local practice, or both. The reasons were not captured sufficiently to separate these explanations. Site variation was substantial, and exploratory analyses did not demonstrate a consistent relationship with enrolment order.^{3,4}

Rescue treatment complicates rather than invalidates the comparison. Conventional platelets given to the cryopreserved group could compensate for inadequate initial treatment and reduce an underlying product difference. Conversely, the decision to rescue is related to bleeding severity and other evolving clinical information. It is not possible to infer the effect of exclusive cryopreserved treatment simply by removing those patients. The reduced per-protocol sample loses randomised comparability, and its confidence interval still crosses the noninferiority boundary.

The primary time window starts at ICU admission even though most patients received their first study unit in theatre. An effect during the operation could therefore be missed by the principal endpoint. This makes the additional intraoperative and early-drain findings relevant, but they are not equally secure. Intraoperative loss was less consistently measured and includes surgical bleeding unrelated to platelet function. Adding the admission drain volume was post hoc. These results broaden the clinical picture without retrospectively redefining the trial's primary success criterion.

The pattern across transfusion and bleeding outcomes is concerning. Higher postoperative loss, more red-cell exposure, more plasma and more cryoprecipitate all point in a similar direction. However, these are correlated consequences in the same patients, not independent replications. Forty-two tertiary outcomes create many opportunities for apparently positive comparisons, and some effects weakened after adjustment. The most defensible interpretation is a coherent signal of less effective haemostasis that reinforces the failure of noninferiority, rather than a catalogue of separately proven harms.

Safety should be described with the same discipline. The absence of a statistically clear difference in a small number of prespecified adverse events does not establish equal safety. Thirteen versus five deaths is a clinically important imbalance, although the confidence interval is wide and does not establish increased mortality. It is equally inappropriate to claim a mortality hazard as proven or to dismiss it because its P value would exceed a conventional threshold. The accompanying editorial was particularly concerned about this distinction.⁴

Competing-risk analysis was appropriate for discharge and extubation, since death prevents these desirable events. A lower discharge incidence can reflect both delayed recovery and more deaths, however, and should not be interpreted as a pure additional number of hospital days caused by the platelet product. The reported medians and the subdistribution hazard ratios answer related but different questions. Both belong within an exploratory assessment rather than a definitive resource-use claim.

Biology remains plausible but unresolved. A lower circulating platelet count does not by itself prove inadequate haemostasis, because activated cells or microparticles might contribute to clotting without persisting in circulation. Conversely, laboratory evidence of increased thrombin generation cannot override greater clinical bleeding. The trial's limited viscoelastic measurements did not resolve the mechanism. The entire product, including freezing, residual dimethyl sulfoxide, plasma resuspension and donor preparation, was randomised as a package.

Applicability is strongest to high-risk adult cardiac surgery in well-resourced centres. The cohort included nonelective and complex operations, but important coagulation disorders and recent thromboembolism were excluded, and most patients came from three hospitals. Results do not directly answer prophylactic transfusion in marrow failure, obstetric haemorrhage or trauma. They also do not establish whether these platelets are beneficial when the true alternative is no platelet product at all.

That final distinction preserves the original clinical problem. Longer storage could still be useful where supply is absent, but usefulness cannot be assumed solely from logistical appeal or from the absence of a definitive toxicity signal. Conversely, a failed comparison with readily available conventional platelets does not prove that all frozen preparations are ineffective. Manufacturing differences and the comparator available in practice remain central to the next clinical question.

Body of Evidence

The early cardiac-surgery study by Khuri and colleagues randomised seventy-three patients and reported less bleeding and postoperative transfusion with cryopreserved platelets despite lower platelet increments. It supported the possibility that laboratory recovery and haemostatic activity might diverge. However, its preparation method differed from CLIP-II, including the timing of dimethyl-sulfoxide removal, and the later report raises concerns about exclusions from the analysed population. It is an early signal rather than a precise contradiction of the larger trial.^{1,5}

CLIP-I randomised 121 high-risk surgical patients, but only forty-one received study platelets: twenty-three cryopreserved and eighteen conventional. Its primary purpose was feasibility and safety. Blood loss and red-cell requirements did not differ clearly, while cryopreserved recipients received more study platelets and more plasma. Small numbers and broad intervals made the absence of an identified major harm insufficient to establish clinical equivalence. CLIP-II supplied the larger effectiveness test that the pilot explicitly required.⁶

The New Zealand pilot similarly randomised eighty-nine patients but transfused only twenty-three. It demonstrated feasibility, with no clear difference in blood loss or recorded safety outcomes, and found a lower next-day platelet concentration with cryopreservation. These observations were useful for planning but could not exclude the clinically important differences subsequently compatible with CLIP-II.⁷

A dose-escalation study in bleeding thrombocytopenic patients gave cryopreserved platelets to twenty-four patients and conventional platelets to four. Bleeding scores

improved in some cryopreserved recipients, and no related thrombotic or serious adverse event was identified. The very small comparator group and different clinical setting limit its ability to answer comparative effectiveness in cardiac surgery. Improvement after transfusion is not equivalent to proof of superiority over another product.⁸

The December 2025 editorial also discussed a conference-reported noninferiority result from CRYPTICS, while noting that the complete data were not then publicly available. That announcement is not treated here as a fully appraised published trial. Resolving apparent differences requires the actual patient populations, manufacturing methods, margins and outcome data, rather than comparing headline conclusions.⁴

Taken together, the earlier studies justified a definitive comparison; they did not guarantee its outcome. CLIP-II failed to establish the required haemostatic noninferiority and produced a consistent, although exploratory, pattern favouring conventional platelets. Its strongest immediate implication is against routine replacement of available liquid-stored platelets with this cryopreserved preparation in cardiac surgery. The separate problem of providing an effective platelet product where none is available remains unanswered.

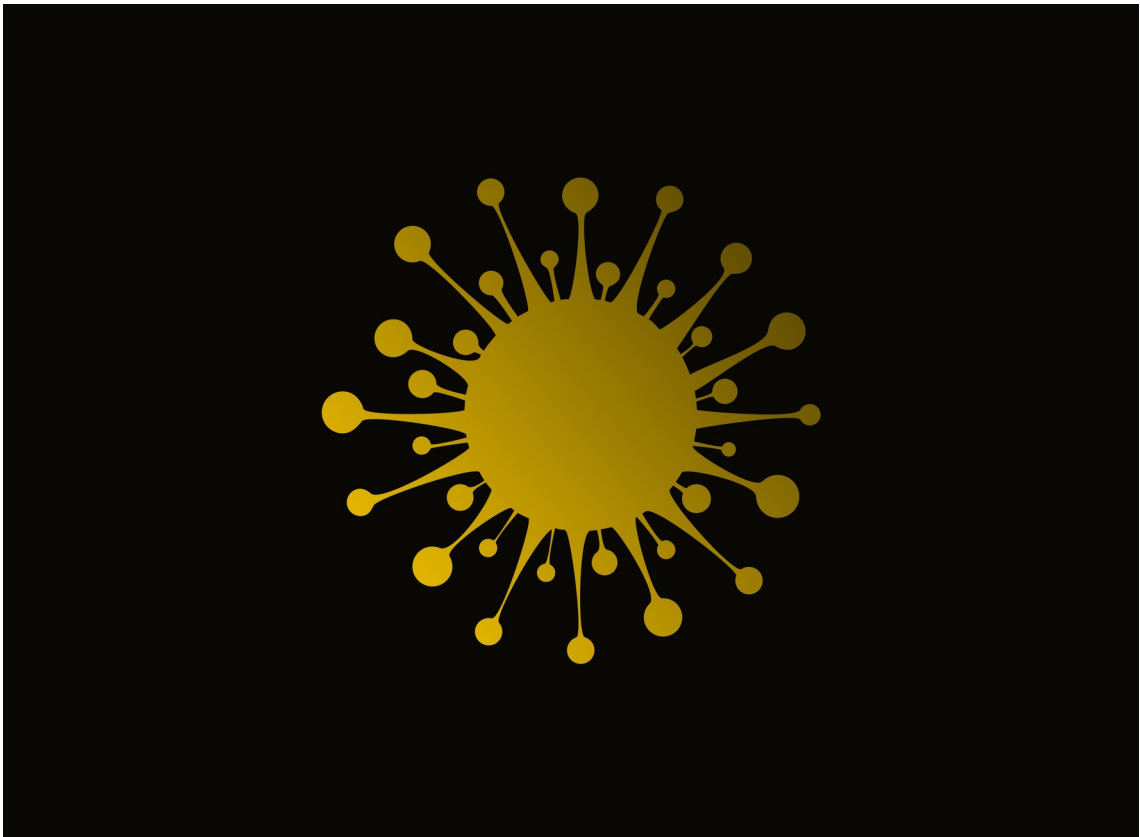
Should cryopreserved platelets routinely replace liquid-stored platelets for cardiac-surgical bleeding?

No. CLIP-II did not establish noninferiority and several bleeding outcomes favoured liquid-stored platelets; benefit where conventional platelets are unavailable remains untested.

References

1. Reade MC, Marks DC, Howe BD, Bailey MJ, Bannon PG, Eastwood GM, et al. Cryopreserved vs Liquid-Stored Platelets for the Treatment of Surgical Bleeding: The CLIP-II Randomized Noninferiority Clinical Trial. JAMA. 2026;335(7):600-608. First published online in 2025.
2. CLIP-II investigators. Investigator's brochure: cryopreserved apheresis platelets reconstituted in thawed clinical fresh frozen plasma. Version 2.0, 20 September 2022.
3. Reade MC, Marks DC, Howe BD, et al. Supplementary methods and results to the CLIP-II trial. JAMA. Published online 8 December 2025. doi:10.1001/jama.2025.23355.

4. Neal MD, Spinella PC, Kornblith LZ. An Oasis in the Platelet Desert? JAMA. 2026;335(7):583-584. First published online in 2025.
5. Khuri SF, Healey N, MacGregor H, Barnard MR, Szymanski IO, Birjiniuk V, et al. Comparison of the effects of transfusions of cryopreserved and liquid-preserved platelets on hemostasis and blood loss after cardiopulmonary bypass. J Thorac Cardiovasc Surg. 1999;117(1):172-83; discussion 183-4.
6. Reade MC, Marks DC, Bellomo R, Deans R, Faulke DJ, Fraser JF, et al. A randomized, controlled pilot clinical trial of cryopreserved platelets for perioperative surgical bleeding: the CLIP-I trial (Editorial, p. 2759). Transfusion. 2019;59(9):2794-2804.
7. McGuinness S, Charlewood R, Gilder E, Parke R, Hayes K, Morley S, et al. A pilot randomized clinical trial of cryopreserved versus liquid-stored platelet transfusion for bleeding in cardiac surgery: The cryopreserved versus liquid platelet-New Zealand pilot trial. Vox Sang. 2022;117(3):337-345.
8. Slichter SJ, Dumont LJ, Cancelas JA, Jones M, Gernsheimer TB, Szczepiorkowski ZM, et al. Safety and efficacy of cryopreserved platelets in bleeding patients with thrombocytopenia. Transfusion. 2018;58(9):2129-2138.



Sepsis Trials

OPTPRESS

Endo A, Yamakawa K, Tagami T, Umemura Y, Wada T, Yamamoto R, et al. Efficacy of targeting high mean arterial pressure for older patients with septic shock (OPTPRESS): a multicentre, pragmatic, open-label, randomised controlled trial. *Intensive Care Med.* 2025;51(5):883-892.

Introduction

A blood pressure target in septic shock is also a prescription for treatment. Raising mean arterial pressure may improve the driving pressure for organ perfusion, but usually requires additional vasopressors, with effects on cardiac workload, regional circulation and arrhythmia risk. The clinical question is therefore not whether a higher pressure looks more reassuring, but whether the treatment needed to achieve it improves outcomes.

Older patients create particular uncertainty. Longstanding hypertension may shift autoregulation towards higher pressures, while frailty, vascular disease and limited cardiac reserve may increase the hazards of vasoconstriction. These considerations point in different directions. A history of hypertension is an imperfect proxy for the pressure an individual patient needs during acute infection, and normalising a previously high blood pressure may require considerable pharmacological support.

SEPSISPAM found no mortality advantage from targeting 80–85 rather than 65–70 mmHg, although a renal signal in patients with chronic hypertension encouraged interest in individualisation. The 65 trial subsequently examined a lower target in older patients with vasodilatory hypotension, without demonstrating a significant reduction in its primary mortality comparison. Neither directly settled the question in an older Japanese population treated very early in septic shock with a strategy incorporating early vasopressin. OPTPRESS tested whether that combination made a high pressure target beneficial.^{1,2,3}

Synopsis

OPTPRESS was a pragmatic, open-label randomised trial at 29 hospitals in Japan. Patients were recruited from July 2021 to December 2023. The Japan Society for the Promotion of Science funded the study and had no role in its conduct, analysis or publication. Written consent was obtained from patients or relatives, with subsequent consent permitted when emergency treatment preceded enrolment formalities. Statisticians remained masked to allocation during analysis.^{3,4}

Eligible patients were at least 65 years old, had septic shock defined using Sepsis-3 criteria and required intensive care. Patients who had already received vasopressors for three hours or longer were excluded. The amount of fluid required before diagnosing shock was left to clinical judgement rather than a mandatory uniform weight-based bolus. This was therefore an early haemodynamic strategy in established septic shock, not a late intervention for persistent hypotension after prolonged treatment.

A central online system allocated patients equally to mean arterial pressure targets of 80–85 or 65–70 mmHg. Randomisation used blocks of ten, with stratification by age below or above 80 years and the presence of known chronic hypertension. Hypertension was identified from the history available at enrolment, including previous diagnosis or antihypertensive treatment. Treating clinicians knew allocation, but the block size was not disclosed to the study teams.

The assigned target applied for 72 hours or until vasopressors were no longer required. Upper-arm non-invasive measurement was preferred in principle. In the high-target group, a potentially vasopressor-related adverse event could trigger reduction of the target to 65 mmHg. After 72 hours, pressure targets were left to the treating team.

The same vasopressor sequence applied in both arms. Vasopressin was to be added when noradrenaline reached 0.1 micrograms/kg/min and titrated up to 0.04 units/min. If the target remained unmet, clinicians could increase noradrenaline or use additional treatments, including another vasopressor, dobutamine or hydrocortisone. Fluid administration, vasopressor tapering, antimicrobial choice, ventilation and renal replacement therapy otherwise remained clinician-directed. Thus, randomisation changed the target within a shared drug strategy; it did not separately randomise vasopressin exposure.

The primary outcome was all-cause mortality 90 days after randomisation. Secondary outcomes included 28-day mortality, deaths attributed to sepsis, lactate change, organ support-free days and adverse events. Patients dying within 28 days were assigned zero organ support-free days. A composite of major adverse kidney events at hospital

discharge was collected retrospectively and was not a prespecified endpoint. It included death, ongoing renal replacement and persistent loss of kidney function.

The investigators expected 45% mortality in controls and sought an absolute reduction of ten percentage points. With 80% power and a two-sided 5% significance level, 752 analysable patients were required; allowing for attrition gave a recruitment target of 836. The principal comparison used Fisher's exact test and reported the absolute risk difference. Prespecified sensitivity analyses adjusted for randomisation strata, while a further adjustment for potential baseline imbalance was post hoc.

An interim analysis was planned after 300 patients had completed 90-day follow-up. Low Bayesian predictive power for demonstrating benefit at the final sample size could prompt discontinuation. There was no efficacy stopping decision requiring adjustment of the final significance threshold. Data-entry and validation delays meant the interim review took approximately eight months after the 300th enrolment, during which recruitment continued. The monitoring committee ultimately recommended stopping because the strategy met the ineffectiveness criterion and the observed direction suggested harm.

Of 561 screened patients, 518 were randomised, 258 to the high target and 260 to the standard target. One erroneous registration in each group was excluded, leaving 257 and 259 for the primary analysis. Median age was 78 years overall. Baseline mean arterial pressure was approximately 56–58 mmHg, lactate approximately 4 mmol/L and Sequential Organ Failure Assessment score approximately nine to ten. Abdominal, urinary and pulmonary infections predominated. About 53% had a recorded history of hypertension. The 66.9% hypertension prevalence mentioned in the trial rationale came from a national survey and does not describe the enrolled cohort.

Randomisation occurred a median of one hour after starting noradrenaline. Median time to first reaching the assigned target was three hours in the high-target group and one hour in controls. Sixteen versus five patients died before reaching their respective targets. Blood pressure separated over the intervention, although by less than the nominal 15-mmHg target difference. At 24 hours, mean pressures in the supplement were 82 versus 75 mmHg; by 72 hours they were 85 versus 83 mmHg, as more patients recovered and came off vasopressors.

Median vasopressor duration was 65 versus 42 hours. Over the first 72 hours, vasopressin exposure was 34.0 versus 11.8 units, and noradrenaline exposure was also greater with the high target. The main text gives a high-target noradrenaline median of 15.6 mg, whereas the supplement gives 15.4 mg, against 9.6 mg in controls. Fluid

volumes were similar, approximately seven litres over 72 hours. Only two high-target patients had their target reduced because of reported adverse events.^{3,4}

At 90 days, 101 of 257 high-target patients, 39.3%, and 74 of 259 controls, 28.6%, had died. The absolute risk difference was 10.7 percentage points, with a 95% confidence interval of 2.6–18.9 and $P=0.012$. At 28 days, mortality was 30.4% versus 21.6%, a difference of 8.7 percentage points with a 95% confidence interval of 1.2–16.3. The stratification-adjusted and per-protocol analyses supported the direction of the primary comparison.

Mean renal replacement therapy-free days were lower by 2.5 days, ventilator-free days by 2.7 days and catecholamine-free days by 2.4 days with the high target. Their confidence intervals excluded zero, but these were secondary analyses without a reported multiplicity-controlled confirmatory hierarchy. Arrhythmias were recorded in 34 versus 29 patients, ischaemic events in 23 versus 14 and haemorrhagic events in 14 versus eight. None individually established a statistically clear difference. Exploratory analyses did not identify a mortality-benefit subgroup; the mortality interaction with hypertension was not significant.

Critique

The clinically important finding is increased mortality after allocation to the higher pressure strategy. It is supported by the direction of 28-day mortality and organ support-free outcomes, and persists in reported sensitivity analyses. This is stronger evidence against routine use of the tested strategy than a neutral superiority trial would provide. Nevertheless, the point estimate should not be treated as a precise universal measure of the harm from every pressure above 70 mmHg.

The randomised intervention was a treatment policy, not achieved pressure in isolation. Patients allocated to 80–85 mmHg received vasopressors for longer and accumulated more of both noradrenaline and vasopressin. These differences are expected consequences of the target and belong within its causal effect. They are not baseline confounding that should simply be adjusted away. Equally, the trial cannot determine whether the adverse effect came predominantly from vasopressin, catecholamine exposure, prolonged treatment, altered regional flow or another downstream consequence.

This distinction underlies the published exchange about early vasopressin. Correspondents questioned whether the greater vasopressin exposure explained the mortality difference. The investigators emphasised that both arms used the same

sequence and that vasopressin was capped while noradrenaline could rise further. Detailed dose trajectories and invasive haemodynamic data were unavailable. Both points can be accepted: the target strategy produced the observed outcome, while a different drug strategy might produce a different effect. Neither the claim that vasopressin caused the deaths nor the claim that it played no role is established.^{5,6}

Early stopping also affects interpretation. The planned benefit was based on 45% control mortality, substantially above the observed 28.6%. Recruitment stopped at 518 rather than 836 patients after an interim analysis intended to assess the prospects of benefit. The final result points towards harm, but an early-stopped estimate may overstate the magnitude of the underlying difference. Its confidence interval still spans a wide range of excess mortality. The delayed interim process is a practical concern for trial monitoring; it does not justify excluding patients recruited while the review was pending.

Patients who died before attaining their assigned target must remain in the randomised comparison. More such deaths occurred in the high-target arm, and correspondence suggested examining an analysis excluding them. Achieving a target is a post-randomisation event influenced by disease severity and treatment. Removing early failures would preferentially select survivors and answer a different, biased question. The investigators correctly retained them when evaluating the policy of pursuing the target.^{7,8}

The non-invasive measurement strategy introduces uncertainty about how closely recorded pressure tracked intra-arterial pressure in individual patients. Because it applied to both groups, it does not automatically invalidate randomisation. However, measurement error need not have an identical treatment consequence at two different thresholds: the same discrepancy can trigger different dose changes. The trial supports the tested monitoring-and-treatment pathway, while its exact translation to continuous invasive monitoring remains a question of applicability.

Achieved pressures must also be interpreted in relation to recovery. Control pressures rose above the assigned range as vasopressors were discontinued. This is not necessarily non-adherence: a patient whose pressure rises spontaneously should not receive treatment to lower it into a vasopressor target. Conversely, patients with severe shock may fail to reach a high target despite substantial exposure. Observational comparisons of attained pressures would consequently mix treatment response, recovery and severity. The assigned strategy remains the sounder basis for inference.

Hypertension does not provide a demonstrated exception. It was determined pragmatically from information available during an emergency, and subgroup denominators differ from the later baseline-history totals. The exploratory mortality interaction was not significant. A significant interaction for retrospectively collected major adverse kidney events was highlighted in the authors' reply, but that composite includes death and showed no clear benefit within the hypertensive subgroup. It cannot support routinely raising the target in all patients with a history of hypertension.

The organ support-free results are clinically relevant but not independent proof of organ-specific toxicity. Death assigns zero days, so increased mortality necessarily worsens these distributions. Similarly, a composite kidney outcome that includes death cannot establish that the intervention directly injured the kidney. The lack of benefit across these measures supports the overall clinical conclusion without identifying a single mechanism.

Safety reporting is less definitive than mortality ascertainment. The table's footnote specifies a smaller treated safety population than the denominators implied by several printed percentages. Non-serious events were recorded retrospectively, detailed timing was incomplete and later events were not comprehensively characterised. Counts can be described, but these data cannot reliably explain the mortality excess or exclude a particular adverse mechanism. Attribution of additional deaths to sepsis by unblinded treating clinicians is especially weak evidence for a proposed immunological drug effect.

Finally, the study concerns older, predominantly lean Japanese patients treated early in septic shock. It does not test a brief pressure challenge for persistent hypoperfusion, a personalised target based on an observed response, or treatment of a patient with another compelling perfusion requirement. Its practical message is narrower and strong: age and a possible history of hypertension are insufficient reasons to routinely impose 80–85 mmHg using this vasopressor strategy.

Body of Evidence

SEPSISPAM enrolled 776 patients and compared the same nominal targets, 80–85 and 65–70 mmHg. Twenty-eight-day mortality was 36.6% versus 34.0%, with a hazard ratio of 1.07 and a 95% confidence interval of 0.84–1.38. Higher targeting increased new atrial fibrillation. Reduced renal replacement in the chronic-hypertension subgroup did not translate into improved survival. OPTPRESS therefore adds a mortality harm signal to a prior trial that had already failed to establish an overall survival benefit.¹

The OVATION pilot examined 60–65 versus 75–80 mmHg in 118 patients with presumed vasodilatory shock. It achieved a mean separation of approximately 9 mmHg and demonstrated the feasibility of testing different targets, rather than providing a definitive mortality comparison. Its small older-age subgroup suggested possible benefit from lower targeting, but could only generate a hypothesis. It helped motivate larger studies while illustrating why a subgroup signal should not become a general treatment rule.⁹

The 65 trial randomised 2,600 patients aged at least 65 years with vasodilatory hypotension to a target of 60–65 mmHg or usual care. Among 2,463 with primary outcome data, 90-day mortality was 41.0% versus 43.8%, an absolute difference of –2.85 percentage points with a 95% confidence interval of –6.75 to 1.05. Vasopressor exposure fell, but the primary unadjusted mortality comparison was not statistically significant. Its broader population and different comparator distinguish it from OPTPRESS; it does not establish that every older patient should receive the lowest tested pressure.²

VASST addressed drug selection rather than pressure selection. In 778 treated patients with septic shock, adjunctive low-dose vasopressin did not significantly reduce 28-day mortality compared with additional noradrenaline: 35.4% versus 39.3%. Its vasopressin range was 0.01–0.03 units/min, and the background strategy differed from OPTPRESS. This supports neither attributing OPTPRESS entirely to vasopressin nor assuming that substituting vasopressin makes aggressive pressure targeting benign. Drug and target questions require distinct comparisons.¹⁰

Across this evidence, routine pursuit of a higher numerical pressure has not reliably improved patient-centred outcomes. OPTPRESS makes that caution more concrete in early septic shock among older patients. The remaining question is how to identify an individual whose perfusion improves sufficiently with a higher pressure to justify the additional treatment, rather than whether a high target should be applied simply because the patient is older.

Should we routinely target a mean arterial pressure of 80–85 mmHg in older patients with septic shock?

No. In OPTPRESS, this strategy increased mortality despite early vasopressin use, and chronic hypertension did not identify a subgroup proven to benefit.

References

1. Asfar P, Meziani F, Hamel JF, Grelon F, Megarbane B, Anguel N, et al. High versus low blood-pressure target in patients with septic shock. *N Engl J Med*. 2014;370(17):1583-1593.
2. Lamontagne F, Richards-Belle A, Thomas K, Harrison DA, Sadique MZ, Grieve RD, et al. Effect of Reduced Exposure to Vasopressors on 90-Day Mortality in Older Critically Ill Patients With Vasodilatory Hypotension: A Randomized Clinical Trial. *JAMA*. 2020;323(10):938-949.
3. Endo A, Yamakawa K, Tagami T, Umemura Y, Wada T, Yamamoto R, et al. Efficacy of targeting high mean arterial pressure for older patients with septic shock (OPTPRESS): a multicentre, pragmatic, open-label, randomised controlled trial. *Intensive Care Med*. 2025;51(5):883-892.
4. Endo A, Yamakawa K, Tagami T, Umemura Y, Wada T, Yamamoto R, et al. Electronic supplementary material 3 to Efficacy of targeting high mean arterial pressure for older patients with septic shock (OPTPRESS). *Intensive Care Med*. 2025.
5. Shi R, Wu J, Teboul JL. OPTPRESS or OPTPRESSIN? *Intensive Care Med*. 2025;51:1561-1562.
6. Endo A, Yamakawa K, Tagami T, Umemura Y. OPTPRESS or OPTPRESSIN? Authors' reply. *Intensive Care Med*. 2025;51:1563-1564.
7. Tiruvoipati R, Kansal A, Panwar R. High mean arterial pressure (MAP) targets in septic shock: one size does not fit all. *Intensive Care Med*. 2025;51:1557-1558.
8. Endo A, Yamakawa K, Tagami T, Umemura Y. High mean arterial pressure (MAP) targets in septic shock: one size does not fit all. Authors' reply. *Intensive Care Med*. 2025;51:1559-1560.
9. Lamontagne F, Meade MO, Hébert PC, Asfar P, Lauzier F, Seely AJE, et al. Higher versus lower blood pressure targets for vasopressor therapy in shock: a multicentre pilot randomized controlled trial. *Intensive Care Med*. 2016;42(4):542-550.
10. Russell JA, Walley KR, Singer J, Gordon AC, Hébert PC, Cooper DJ, et al. Vasopressin versus norepinephrine infusion in patients with septic shock. *N Engl J Med*. 2008;358(9):877-887.

ANDROMEDA-SHOCK-2

The ANDROMEDA-SHOCK-2 Investigators for the ANDROMEDA Research Network, Spanish Society of Anesthesiology, Reanimation and Pain Therapy (SEDAR), and Latin American Intensive Care Network (LIVEN). Personalized Hemodynamic Resuscitation Targeting Capillary Refill Time in Early Septic Shock: The ANDROMEDA-SHOCK-2 Randomized Clinical Trial. JAMA. 2025;334(22):1988-1999.

Introduction

Septic shock is treated with interventions that can improve the circulation and also cause harm. Fluid may increase stroke volume in a responsive patient but accumulate when further loading is ineffective. Vasopressors can restore perfusion pressure while increasing cardiac afterload and altering regional blood flow. Inotropes may improve output at the cost of tachycardia or hypotension. The practical challenge is deciding which intervention a particular patient needs, whether it is working and when to stop.

Capillary refill time offers a rapid bedside indication of peripheral perfusion. It can be measured repeatedly without waiting for laboratory results, whereas lactate may remain elevated because of processes other than ongoing flow-limited tissue hypoxia. A normal refill time might therefore help prevent unnecessary additional resuscitation. An abnormal value, however, does not identify whether the cause is inadequate preload, vasoplegia, ventricular dysfunction or a persisting microcirculatory disturbance.

The original ANDROMEDA-SHOCK trial compared resuscitation targeting capillary refill with a lactate-targeted strategy. Its primary frequentist mortality result remained inconclusive, although several findings encouraged further evaluation. ANDROMEDA-SHOCK-2 developed a broader approach: use capillary refill as the treatment target, combine it with simple haemodynamic assessment and fluid responsiveness, and escalate selectively through echocardiography and short therapeutic tests. The question was whether this whole strategy improved outcomes compared with contemporary usual care.^{1,2}

Synopsis

ANDROMEDA-SHOCK-2 was an investigator-led, open-label randomised trial at 86 intensive care units in 19 countries across the Americas, Europe and Asia. Recruitment ran from March 2022 to April 2025, with follow-up completed in July. Academic institutions and public research programmes provided logistical and financial support. An independent monitoring board oversaw the trial and conducted two interim reviews. Consent arrangements followed local requirements.^{2,3}

Adults were eligible within four hours of meeting the study definition of septic shock: suspected or confirmed infection, lactate of at least 2 mmol/L and noradrenaline required to maintain mean arterial pressure of at least 65 mmHg after at least one litre of intravenous fluid. Patients could be identified in intensive care, the emergency department, operating theatre or hospital wards.

Important exclusions included refractory shock, anticipated surgery or dialysis during the next six hours, severe acute respiratory distress syndrome, active bleeding, advanced liver disease and limited life expectancy. Patients in whom refill time could not be assessed reliably were also excluded. The screening diagram identifies peripheral vascular disease, hypothermia, darker skin tone and rheumatological disease among potential reasons for unreliable assessment. These exclusions are relevant to implementation of a visually assessed perfusion target.

Central concealed web randomisation assigned patients equally to capillary refill-targeted personalised haemodynamic resuscitation or usual care, using variable blocks and stratification by centre. Patients and clinicians were not masked. Intervention clinicians completed mandatory training with videos, manuals, case exercises and virtual sessions. Adherence was audited with prompt feedback. Control patients were managed by their attending intensivists.

The intervention lasted six hours. Refill time was measured after applying firm pressure with a glass slide to the distal finger for ten seconds, then timing return of normal colour. More than three seconds was abnormal. Patients whose refill time was normal underwent hourly reassessment without further algorithm-directed resuscitation unless it became abnormal. Standard care goals continued to apply.

For abnormal refill, the first tier used pulse pressure and diastolic pressure to direct treatment. A pulse pressure below 40 mmHg prompted fluid-responsiveness assessment, followed, if positive, by a 500-mL bolus over 30 minutes and reassessment. A second bolus could follow if the patient remained responsive with abnormal refill. With

pulse pressure at least 40 mmHg and diastolic pressure below 50 mmHg, noradrenaline was adjusted to raise diastolic pressure to at least 50 mmHg.

Persistent abnormal refill prompted the second tier, starting with basic echocardiography. Identified ventricular dysfunction was managed using general recommendations rather than one mandatory treatment. Fluid responsiveness was then reassessed, with additional boluses only while responsiveness persisted and safety limits allowed. In patients with chronic hypertension who remained abnormal, a one-hour pressure test raised mean arterial pressure to 80–85 mmHg. That target was continued only if refill normalised; otherwise the previous noradrenaline dose was restored.

A subsequent one-hour dobutamine test used 5 micrograms/kg/min, again continuing only if refill normalised. Patients already receiving dobutamine for cardiac dysfunction did not repeat this step. Clinicians could override the algorithm for safety. Once refill normalised, further resuscitative interventions stopped and hourly reassessment continued.

Usual care followed local practice or international guidance. Echocardiography and fluid-responsiveness testing were permitted but not mandated, and refill measurements were requested only at baseline and six hours. After the intervention period, both groups returned to attending-team management. The comparison therefore included differences in the organisation and frequency of reassessment as well as the treatment algorithm.

The primary outcome was a hierarchical composite through day 28, analysed with a win ratio. Each intervention patient was compared with control patients within one of two baseline APACHE II severity strata. Survival was considered first. If one patient survived and the other died, the survivor won; if both died, the pair tied without further comparison. If both survived, shorter time to final cessation of all vital support determined the winner, followed, if still tied, by shorter hospital stay. Vital support included vasoactive drugs, mechanical ventilation and kidney replacement therapy, with resumed treatment counted until its final cessation.

The planned sample of 1,500 gave 88% power under assumptions of mortality falling from 39% to 33%, mean support duration among survivors falling from 5.6 to 4.3 days, and hospital stay falling from 15.6 to 14 days. The primary analysis used a stratified win ratio with bootstrap confidence intervals. Secondary outcomes followed a gatekeeping sequence: mortality, then vital support-free days, then hospital stay. Testing stopped when a preceding comparison failed to establish significance. Tertiary and subgroup analyses were exploratory.

Of 4,553 patients screened, 1,501 were randomised: 744 to the intervention and 757 to usual care. Twenty withdrew consent and fourteen had been incorrectly enrolled. The analysis included 720 and 747, respectively. This was an eligibility-and-consent-restricted analysis rather than an unrestricted analysis of every randomised patient. Median age was approximately 65–66 years, median Sequential Organ Failure Assessment score eight and median lactate 3.6–3.7 mmol/L. Nearly half had an abdominal source and approximately half received invasive ventilation. Randomisation occurred about two hours after meeting shock criteria, after a median 1.5 litres of fluid.

In the intervention group, 262 patients, 36.4%, maintained normal refill throughout six hours and received no further algorithm-directed haemodynamic intervention. Approximately 35% progressed to the second tier. In controls, 62% underwent fluid-responsiveness assessment and 68% basic echocardiography. Usual care was therefore active and often used similar diagnostic tools.

Mean resuscitation fluid during the intervention was 595 versus 847 mL, with a reported difference of –251 mL and a 95% confidence interval of –316 to –187. Dobutamine use at six hours was 12.3% versus 5.3%, while mean arterial pressure was similar, about 74 mmHg. Refill normalised in 85.9% versus 61.7%. Protocol deviations occurred in 15% and violations in 6% of intervention patients. Some haemodynamic tests were withheld for safety, and one dobutamine test stopped for reversible hypotension.

For the primary outcome, intervention patients won 131,131 pairwise comparisons, 48.9%, and controls won 112,787, 42.1%; 24,151 comparisons tied. The stratified win ratio was 1.16, with a 95% confidence interval of 1.02–1.33 and $P=0.04$. Most of the separation arose from duration of vital support rather than death or hospital stay.^{2,3}

Mortality at 28 days was 191 of 720, 26.5%, versus 199 of 747, 26.6%, with an adjusted hazard ratio of 0.99 and a 95% confidence interval of 0.81–1.21. Because mortality did not pass the first secondary gate, later secondary outcomes did not receive confirmatory significance claims. Mean vital support-free days were 16.5 versus 15.4, with a proportional odds ratio of 1.28 and a 95% confidence interval of 1.06–1.54. Median duration of vital support was three versus four days. At 90 days, mortality was 32.1% versus 33.2%, with a hazard ratio of 0.93 and a 95% confidence interval of 0.78–1.11. No convincing subgroup interaction or protocol-related serious adverse reaction was identified.

Critique

ANDROMEDA-SHOCK-2 established superiority for its prespecified hierarchical composite. It did not establish a mortality benefit. That distinction is not a reason to dismiss shorter organ support: liberation from ventilation, vasoactive therapy or dialysis matters to patients and health systems. It is a reason to describe the result according to what actually generated the advantage.

The win ratio requires particular care. A value of 1.16 means that favourable pairwise comparisons outnumbered unfavourable comparisons within the chosen hierarchy. It does not mean a 16% reduction in mortality, a 16% increase in individual survival probability or that 16% of patients avoided organ support. The hundreds of thousands of pairs are comparisons among the same 1,467 patients, not independent participants. Bootstrap uncertainty must respect the patient-level data, and the reported confidence interval remains relatively close to the null.

Placing death first gives it priority within each pair but does not guarantee that mortality dominates the overall result. With almost identical death rates and lower mortality than anticipated, many comparisons advanced to support duration. Those survivor comparisons supplied most of the advantage. Both-death pairs became early ties, so the analysis did not reward merely postponing death. These are defensible choices, but the resulting estimand is a particular ordering of clinical outcomes, not a single intuitive quantity such as days alive at home.

The secondary testing hierarchy was appropriately conservative. The non-significant mortality comparison closed the gate to confirmatory testing of later secondary outcomes. Consequently, the confidence interval excluding one for support-free days is supportive description, not a second independently confirmed positive endpoint. Conversely, failure at the secondary mortality gate does not invalidate the already-positive primary composite. The two analyses answer related but different questions.

The central vulnerability is that the component driving benefit depended on unblinded clinical decisions. Withdrawal of vasopressors, extubation and discontinuation of kidney support followed local practice without independent adjudication. Teams who observed normalisation of a target they had actively pursued might be more confident about withdrawing support. This could represent appropriate earlier liberation, a behavioural effect of structured reassessment or biased outcome management; the trial cannot fully separate them. Mortality, the least discretionary component, did not improve.

The intervention was also a bundle. It combined frequent bedside review, a stopping rule, fluid-responsiveness assessment, pressure-based branching, echocardiography and selective drug tests. Randomisation supports the effect of offering that package. It does not validate each pulse-pressure threshold, prove that a diastolic pressure of 50 mmHg is optimal or establish dobutamine as beneficial in everyone with delayed refill. The accompanying editorial emphasised both the additional clinical attention and the inability to identify which elements were responsible.⁴

Avoidance of unnecessary treatment may be as important as escalation. More than a third of intervention patients required no additional algorithm-directed resuscitation because refill remained normal. The overall fluid difference was modest, about 250 mL, but an average can combine withholding fluid in one patient with appropriate administration in another. The trial is therefore not adequately described as a restrictive-fluid trial. Nor can post-randomisation response groups identify who benefited causally: patients whose refill normalised after a treatment may already have had a better prognosis.

Control care strengthens relevance while complicating mechanistic attribution. Fluid-responsiveness assessment and echocardiography were common in controls, and their mortality was lower than predicted. The intervention was not compared with an absence of haemodynamic assessment. Training at participating centres could also influence control care, potentially reducing separation, although that possibility cannot quantify how large an effect would occur elsewhere.

The pressure test is materially different from routinely pursuing a high target. It was reserved for selected chronically hypertensive patients with persisting abnormal refill, lasted one hour and was reversed if the target response failed. Thus, the positive composite does not contradict trials cautioning against sustained high-pressure policies. It also cannot establish that this particular test contributed benefit: relatively few patients reached that step, and it was not separately randomised.

Selection limits the breadth of implementation. Patients with refractory shock, imminent surgery or dialysis, severe respiratory failure and unreliable refill assessment were excluded. A visually assessed endpoint may be more difficult in darker skin, peripheral vascular disease, hypothermia or poor lighting. Standardisation and training can reduce variability but do not eliminate these practical limitations. Normal finger refill should not be assumed to exclude every form of clinically important organ hypoperfusion.

The multinational setting is an important strength, but the treatment was delivered by trained operators with active adherence review. A six-hour sequence of assessments is a

clinical service, not merely an instruction to record refill time on a chart. Reproducing the tested strategy requires staff availability, reliable fluid-responsiveness assessment and access to basic echocardiography for persistent abnormalities. The benefit observed under those conditions cannot automatically be assigned to a simplified version from which inconvenient components are removed.

Post-randomisation exclusions were few, but more occurred in the intervention arm: 24 versus ten. They were attributed to consent withdrawal and mistaken eligibility. The retained sample had complete primary follow-up, which is reassuring, yet exclusion after allocation can weaken the protection of randomisation if related to prognosis or treatment. The published result should be presented with its actual analysis population rather than described as including all 1,501 participants.

The trial supports a structured, responsive approach to early resuscitation with capillary refill as a useful bedside target. Its strongest practical implication is to reassess the need for each additional intervention and stop when the targeted abnormality resolves. The evidence is less secure for claiming that this precise sequence is the uniquely optimal algorithm or that any individual component improves survival.

Body of Evidence

The original ANDROMEDA-SHOCK trial randomised 424 patients to capillary refill- or lactate-targeted resuscitation. Twenty-eight-day mortality was 34.9% versus 43.4%, with a hazard ratio of 0.75 and a 95% confidence interval of 0.55–1.02; $P=0.06$. Organ dysfunction at 72 hours favoured refill targeting. This was promising but did not establish superiority for the primary mortality endpoint. The second trial changes both the intervention and comparator, so it is an extension of the clinical programme rather than an exact replication.¹

The PRISM patient-level meta-analysis combined three modern trials of early goal-directed therapy, involving 3,723 patients. Ninety-day mortality was similar with the structured protocol and usual care, while intensive treatment and costs increased. It showed that a plausible multicomponent resuscitation package must demonstrate benefit against contemporary care. ANDROMEDA-SHOCK-2 uses a different target and places more emphasis on stopping or reversing interventions according to response.⁵

CLASSIC compared restricted with standard intravenous fluid after initial resuscitation in 1,554 intensive care patients with septic shock. Ninety-day mortality was 42.3% versus 42.1%, despite substantially less fluid with restriction. CLOVERS compared an early vasopressor-prioritising restrictive strategy with a more liberal fluid strategy in 1,563

patients with sepsis-induced hypotension; mortality before discharge home by day 90 was 14.0% versus 14.9%. Neither found a clear survival advantage. Their principal lesson here is that selecting a general fluid-volume policy differs from using serial physiological response to decide whether an individual needs another bolus.^{6,7}

The smaller FRESH trial randomised 150 patients and used passive leg raising with stroke-volume assessment to guide fluid and vasopressor decisions. Its primary analysis in 124 eligible patients reported lower fluid balance and encouraging organ-support findings, but was not a definitive mortality trial. It supports the feasibility of making responsiveness part of treatment selection, while leaving substantial uncertainty about the best complete resuscitation pathway.⁸

ANDROMEDA-SHOCK-2 adds a positive primary composite to this mixed evidence base, chiefly through shorter vital support. It strengthens the case for capillary refill within structured bedside reassessment, while preserving the need to distinguish a useful clinical signal, an effective package of care and a separately proven treatment component.

Should we use capillary refill-targeted personalised resuscitation in early septic shock?

Consider this structured approach where the necessary assessment and training are available. It improved the hierarchical outcome mainly through shorter vital support, without demonstrating a mortality benefit.

References

1. Hernández G, Ospina-Tascón GA, Damiani LP, Estenssoro E, Dubin A, Hurtado J, et al. Effect of a Resuscitation Strategy Targeting Peripheral Perfusion Status vs Serum Lactate Levels on 28-Day Mortality Among Patients With Septic Shock: The ANDROMEDA-SHOCK Randomized Clinical Trial. *JAMA*. 2019;321(7):654-664.
2. The ANDROMEDA-SHOCK-2 Investigators for the ANDROMEDA Research Network, Spanish Society of Anesthesiology, Reanimation and Pain Therapy (SEDAR), and Latin American Intensive Care Network (LIVEN). Personalized Hemodynamic Resuscitation Targeting Capillary Refill Time in Early Septic Shock: The ANDROMEDA-SHOCK-2 Randomized Clinical Trial. *JAMA*. 2025;334(22):1988-1999.

3. The ANDROMEDA-SHOCK-2 Investigators. Supplement 2 to Personalized Hemodynamic Resuscitation Targeting Capillary Refill Time in Early Septic Shock: The ANDROMEDA-SHOCK-2 Randomized Clinical Trial. JAMA. 2025.
4. Machado FR, Semler MW. Capillary Refill Time in Sepsis-Searching for the Holy Grail. JAMA. 2025;334(22):1983-1985.
5. PRISM Investigators. Early, Goal-Directed Therapy for Septic Shock — A Patient-Level Meta-Analysis. N Engl J Med. 2017;376(23):2223-2234.
6. Meyhoff TS, Hjortrup PB, Wetterslev J, Sivapalan P, Laake JH, Cronhjort M, et al. Restriction of Intravenous Fluid in ICU Patients with Septic Shock. N Engl J Med. 2022;386(26):2459-2470.
7. National Heart, Lung, and Blood Institute Prevention and Early Treatment of Acute Lung Injury Clinical Trials Network. Early Restrictive or Liberal Fluid Management for Sepsis-Induced Hypotension. N Engl J Med. 2023;388(6):499-510.
8. Douglas IS, Alapat PM, Corl KA, Exline MC, Forni LG, Holder AL, et al. Fluid Response Evaluation in Sepsis Hypotension and Shock: A Randomized Clinical Trial. Chest. 2020;158(4):1431-1445.

TARTARE-2S

Pettilä V, Pfortmüller CA, Perner A, Merz TM, Wilkman E, Hästbacka J, et al. Targeted Tissue Perfusion Versus Macrocirculatory-Guided Standard Care in Patients With Septic Shock: A Randomized Clinical Trial—The TARTARE-2S Trial. Crit Care Med. 2026;54(1):24-34. First published online 17 October 2025.

Introduction

A satisfactory arterial pressure does not guarantee satisfactory tissue perfusion. Sepsis can disrupt regional blood flow and microcirculatory function despite apparently adequate systemic measurements. Conversely, a patient may have acceptable clinical perfusion at a pressure below a conventional target. The question is whether treatment should pursue the pressure itself or use pressure as one part of a broader assessment of adequate circulation.

This distinction matters because pressure is often increased with vasopressors. These drugs restore perfusion pressure when it is inadequate, but greater exposure can also impose cardiovascular and peripheral vascular burdens. Raising the target without a demonstrable benefit may therefore add treatment rather than improve resuscitation. The difficult clinical task is to identify when a lower pressure is acceptable and when it threatens organ blood flow.

TARTARE-2S tested a strategy that prioritised tissue perfusion indicators while allowing lower mean arterial pressure. It incorporated capillary refill, mottling, peripheral temperature, lactate and urine output, with higher pressure targets for selected circumstances. This was not simply a comparison of two fixed pressure numbers. It was a phase II evaluation of whether a different set of bedside goals could accelerate shock resolution without an obvious adverse signal.¹

Synopsis

TARTARE-2S was an investigator-initiated, open-label, parallel-group randomised trial in three university intensive care units in Finland, Switzerland and Denmark. Recruitment ran from June 2016 to May 2022, with nine months of interruption during the COVID-19 pandemic. Institutional and foundation funding supported the study. Proxy and deferred consent arrangements permitted early enrolment, with additional independent physician approval where required by national procedures.^{1,2}

Eligible patients were aged 18–80 years and had suspected or confirmed infection, a vasopressor requirement to maintain mean arterial pressure of at least 65 mmHg after initial fluid resuscitation, and arterial lactate of at least 3 mmol/L with clinically suspected hypoperfusion. Initial resuscitation required at least 20 mL/kg of crystalloid within two hours. Eligibility had to be established within four hours of ICU admission, and randomisation occur within four hours of meeting the criteria.

Exclusions included non-septic conditions affecting the central nervous system, including cardiac arrest, myocardial ischaemia, acute pulmonary embolism, severe liver disease, known chronic kidney disease, extracorporeal membrane oxygenation and another likely cause of hyperlactataemia. Patients not considered appropriate for full intensive care, those transferred from another ICU and several additional special populations were also excluded. The high lactate threshold sought a population with established shock rather than transient mild biochemical abnormality.

A concealed computer allocation system randomised patients equally, using variable blocks and stratifying by site and chronic arterial hypertension. Clinicians and patients were not masked. The primary analysis was performed with treatment allocation concealed from the statistician; adverse reaction assessment also used masked assessors.

The targeted tissue perfusion group aimed for capillary refill below three seconds, absent mottling, warm peripheral skin, lactate below 2 mmol/L and urine output of at least 0.5 mL/kg/hour. Mean arterial pressure of 50–65 mmHg was accepted, with the lower boundary acting as a safety limit. Patients with chronic hypertension instead had a target of 65–70 mmHg. Oliguria could prompt a two-hour trial at 75–80 mmHg; if urine output improved, the higher target could continue with repeated reassessment.

Standard care used pressure targets derived from the 2012 Surviving Sepsis Campaign framework: 65–75 mmHg, or 75–80 mmHg with chronic hypertension. The urine output goal was the same. A two-hour trial at 85–90 mmHg was suggested for oliguria, again continued only with a response and reassessment. Fluids treated clinical hypovolaemia, with central venous pressure of 8–12 mmHg used as a safety limit rather than an instruction to give unlimited fluid.

The study provided targets and general treatment suggestions, rather than a mandatory sequence of fluids, vasopressors and inotropes. Clinicians identified haemodynamic problems and selected treatment. Echocardiography and pulmonary artery catheterisation were available at their discretion. Noradrenaline was the preferred

vasopressor, with inotropic treatment for impaired flow. The allocated targets continued until lactate was below 2 mmol/L and vasoactive drugs were no longer required.

Clinical targets were recorded hourly initially and at least two-hourly thereafter. Lactate was measured every two hours until a confirmed value below 2 mmol/L without vasopressors or inotropes, or until 72 hours. The primary outcome was days alive within 30 days with normal lactate and without either vasopressors or inotropes. Each qualifying calendar day contributed one day. Death did not trigger an additional penalty that erased previously accumulated qualifying days.

Secondary outcomes included the components of the primary endpoint, time to lactate normalisation, support-free days, acute kidney injury, renal replacement, cumulative noradrenaline exposure through 96 hours and adverse reactions. Ninety-day mortality was exploratory. The study was described as phase II and was not powered for mortality or other secondary outcomes. Continuous outcomes were compared with a Mann–Whitney test and dichotomous outcomes with Fisher's exact test.

Of 4,343 patients assessed, 1,298 had septic shock and 219 were randomised, 111 to targeted tissue perfusion and 108 to standard care. Nine were subsequently found to meet exclusion criteria and sixteen were removed after consent was refused or withdrawn. The modified intention-to-treat analysis therefore included 194 patients, 97 per group. This represents approximately 17% of screened patients with septic shock before the post-randomisation exclusions.

Mean age was 62 years in both groups, and approximately 40% had chronic hypertension. Mean lactate was 6.2 versus 5.6 mmol/L and mean SOFA scores were 8.9 and 8.6. Mechanical ventilation during the preceding day was reported in 61% versus 73%, and renal replacement in approximately 22% versus 24%. Abdominal infection was the most common specified source. Pulmonary artery catheters were used in 60% of each group on the inclusion day or the following day.

Pressure separation was achieved. Mean arterial pressures over the first 72 hours were reported as 68.5 mmHg with targeted tissue perfusion and 74.8 mmHg with standard care. The tissue perfusion group achieved individual targets with variable frequency: the mean proportion of recordings per patient meeting the lactate target was 31%, capillary refill 40%, absent mottling 71% and warm peripheral temperature 72%. These are repeated-assessment summaries, not percentages of patients who permanently recovered.

The primary outcome was a median of 23 days, with an interquartile range of 10–27, versus 22 days, with an interquartile range of 1–27. Table 2 reports an effect estimate of 0.59 days, with a 95% confidence interval of –3 to 4; $P=0.418$. The study did not demonstrate an improvement. At 30 days, 24 of 97 patients, 24.7%, had died in the targeted group and 27 of 97, 27.8%, in standard care. The risk difference was –3.1 percentage points, with a 95% confidence interval of –16.5 to 10.3.¹

No secondary outcome showed a clear difference. Acute kidney injury occurred in 78 versus 69 patients, and renal replacement associated with acute kidney injury in 35 versus 31. Cumulative fluid exposure and balance were similar, as was total noradrenaline exposure despite different dose patterns over time. Trials of higher pressure for oliguria were common, occurring in 59 versus 57 patients. Glucocorticoids were given to 42 versus 63 patients, 43% versus 65%. Six targeted-group and three standard-care patients had at least one serious adverse reaction.^{1,2}

Critique

The principal finding is absence of a demonstrated improvement in the chosen shock-resolution endpoint. TARTARE-2S does not establish equivalence, and it cannot establish the safety of routinely accepting a pressure of 50 mmHg in septic shock. The mortality interval encompasses substantial benefit and substantial harm. A small superiority trial with few adverse events cannot answer a non-inferiority question merely because its comparisons are not statistically significant.

The achieved pressure deserves more attention than the nominal range. The average in the targeted group was 68.5 mmHg, above the 50–65 mmHg range specified for patients without chronic hypertension. Approximately 40% of participants had higher targets because of hypertension, and more than half underwent temporary pressure trials for oliguria. The intervention therefore did not expose the whole group to sustained pressure near its lower safety limit. Its observed outcomes cannot be used as direct evidence that prolonged pressure at 50–55 mmHg is acceptable.

Above-target pressure also does not always mean excessive vasopressor treatment or failure to follow the strategy. Some readings occurred after vasopressors had been stopped, and others during permitted trials of a higher pressure. The supplementary analysis identifies both explanations, although the precise start and stop times of pressure trials were not documented. This matters because a patient whose pressure spontaneously exceeds a permissive target should not have that pressure actively reduced simply to satisfy the study range.

The trial changed several targets simultaneously. It cannot separate the effects of accepting lower pressure from those of repeated assessment of skin perfusion, urine output and lactate. Nor can it identify which tissue indicator is most useful when they disagree. A normal capillary refill with persistent hyperlactataemia may prompt a different decision from cold extremities with improving lactate, but no fixed hierarchy dictated the response. Clinician judgement was integral to the intervention.

This flexibility increases clinical realism while reducing reproducibility. Two clinicians could observe the same unmet target and choose fluid, more noradrenaline, an inotrope or further diagnostic assessment. Those actions are not physiologically interchangeable. The detailed recording of problems and interventions is a strength, but the result concerns the complete strategy as delivered at these centres, rather than proving or disproving each underlying physiological concept.

The actual contrast with standard care is also uncertain. Tissue perfusion targets were systematically recorded only in the intervention group. Control clinicians could still assess capillary refill, skin temperature and mottling as part of ordinary care. Without comparable recordings, the study cannot establish whether the groups achieved different tissue perfusion states. The most clearly demonstrated separation was arterial pressure.

Target attainment must be interpreted with similar care. Low attainment of a lactate target can reflect persistent illness rather than poor implementation. Conversely, declaring a goal does not show that treatment was effectively directed towards it. The percentages describe repeated observations over a period in which patients improved, died or left the measurement window. They should not be read as simple adherence rates or as proof that a particular proportion of patients received the intended treatment incorrectly.

The primary outcome mixes survival, a biochemical marker and a treatment decision. It is potentially more sensitive to early recovery than mortality alone, but it is not a validated measure with an established minimum clinically important difference. Lactate can remain elevated for reasons other than inadequate perfusion, while vasopressor withdrawal depends partly on the pressure goal assigned to the patient. An intervention that accepts lower pressure could shorten support directly without improving underlying disease; in this trial, even that composite advantage was not demonstrated.

The handling of death is particularly important. Because previously accumulated qualifying days were retained, a patient who recovered temporarily and later died could still contribute a positive number of days. This differs from many organ-support-free-day

outcomes that assign all deaths the worst score. The endpoint therefore does not make survival an overriding priority. Readers should examine mortality alongside the composite rather than assume that a higher score necessarily means a better survival outcome.

Twenty-five of 219 randomised patients were omitted from the principal analysis. Consent refusal or withdrawal creates a legitimate constraint on data use, and newly discovered exclusions may be clinically understandable. Nevertheless, these exclusions occurred after allocation and removed approximately one in nine participants. Calling the analysis modified intention-to-treat describes what happened; it does not preserve all the protection of analysing every randomised patient. The claim of less than 2% missingness among recorded variables should not obscure this larger loss from the randomised cohort.

The small sample leaves uncertainty about secondary outcomes and imbalances. Mechanical ventilation was more frequent in standard care at baseline. During treatment, glucocorticoid use differed substantially, despite general suggestions initially discouraging it. Steroid administration may affect shock duration, but the trial cannot establish why the difference arose or what part it played. It is a post-randomisation co-intervention, not a baseline characteristic that can simply be adjusted away to reveal a pure pressure effect.

Recruitment over almost six years also matters. Sepsis practice evolved, including steroid use and the prominence of capillary refill assessment. Concurrent randomisation helps protect the internal comparison, but the control strategy's 2012 framework became less representative of later practice. The long duration, pandemic interruptions and recruitment from only three specialist units constrain how confidently the delivered strategy can be transferred elsewhere.

Generalisability is further restricted by the exclusions. Patients older than 80 years, those with known chronic kidney disease, active myocardial ischaemia and several conditions affecting lactate or neurological assessment were not represented. These are precisely some of the patients in whom pressure targets are difficult to individualise. The frequent use of pulmonary artery catheters also reflects a level and style of haemodynamic assessment that may differ substantially from many ICUs.

Safety data are reassuring only in a limited sense. No clear excess was detected, but there were six versus three patients with serious adverse reactions, and the renal outcome estimates allow clinically important harm as well as benefit. The strategy's feasibility should be distinguished from a robust demonstration of safety. The

accompanying editorial makes the same central statistical point: wide intervals from a superiority trial cannot justify abandoning established pressure targets.³

Several numerical details warrant editorial clarification. The main table labels the primary estimate a difference in medians but gives 0.59 days, whereas the displayed medians differ by one day and the results text gives one day. The reported mean pressure difference also does not equal the simple subtraction of the displayed group means. The direction and broad uncertainty are clear, but the precise estimands should not be inferred from the labels alone. This chapter retains the table's primary estimate and reports the observed pressure means without silently reconciling the discrepancies.

The appropriate implication is therefore a research question rather than an unrestricted practice change. Integrating tissue perfusion into treatment is clinically attractive, and lower pressure can be delivered within a monitored strategy. Whether this permits a reliably safer reduction in vasopressor exposure, which patients benefit and how low is acceptable remain unresolved.

Body of Evidence

SEPSISPAM found no overall mortality advantage from a target of 80–85 rather than 65–70 mmHg. The 65 trial likewise did not establish a primary mortality benefit from permissive hypotension compared with usual care. Neither tested the complete TARTARE-2S strategy or its lowest permitted pressure.^{4,5}

OPTPRESS adds relevant 2025 evidence from older patients with septic shock. Targeting 80–85 rather than 65–70 mmHg increased vasopressor exposure and was associated with higher 90-day mortality. That result argues against routinely pursuing a high pressure in its population. It does not establish that progressively lower pressures are progressively better, particularly below the range tested. The relationship between inadequate perfusion and excessive treatment need not be linear.⁶

The original ANDROMEDA-SHOCK trial suggested potential benefit from capillary refill-targeted resuscitation but did not meet its primary frequentist mortality threshold. It supported further study of peripheral perfusion without validating every strategy that includes it.⁷

ANDROMEDA-SHOCK-2, published later in 2025, supplies a larger test of a structured capillary refill-based approach. Its personalised six-hour algorithm improved a hierarchical composite, predominantly through shorter vital support, without demonstrating lower mortality. It retained pressure assessment and used physiological

testing to decide whether additional fluid, vasopressor or inotropic treatment helped. It therefore supports structured reassessment rather than a general instruction to tolerate very low pressure whenever one peripheral sign appears normal.⁸

These trials examine related but distinct questions: how high to raise pressure, when to stop adding treatment and how to choose the next intervention. TARTARE-2S contributes early evidence about combining perfusion goals with permissive pressure targets. It does not identify a universal replacement threshold or prove that tissue signs can safely override systemic pressure in every patient.

Can we routinely accept a mean arterial pressure of 50–65 mmHg in septic shock?

No. TARTARE-2S did not improve shock resolution, and its small sample and achieved pressures do not establish the safety of routinely using this low target.

References

1. Pettilä V, Pfortmüller CA, Perner A, Merz TM, Wilkman E, Hästbacka J, et al. Targeted Tissue Perfusion Versus Macrocirculatory-Guided Standard Care in Patients With Septic Shock: A Randomized Clinical Trial-The TARTARE-2S Trial. *Crit Care Med*. 2026;54(1):24-34.
2. Pettilä V, et al. Electronic Supplementary Material to the TARTARE-2S Trial. *Crit Care Med*. 2026.
3. Basmaji J, Chew MS. Between Pressure and Perfusion: The Unresolved Tension in Septic Shock Resuscitation. *Crit Care Med*. 2026;54(1):193-196.
4. Asfar P, Meziani F, Hamel JF, Grelon F, Megarbane B, Anguel N, et al. High versus low blood-pressure target in patients with septic shock. *N Engl J Med*. 2014;370(17):1583-1593.
5. Lamontagne F, Richards-Belle A, Thomas K, Harrison DA, Sadique MZ, Grieve RD, et al. Effect of Reduced Exposure to Vasopressors on 90-Day Mortality in Older Critically Ill Patients With Vasodilatory Hypotension: A Randomized Clinical Trial. *JAMA*. 2020;323(10):938-949.

6. Endo A, Yamakawa K, Tagami T, Umemura Y, Wada T, Yamamoto R, et al. Efficacy of targeting high mean arterial pressure for older patients with septic shock (OPTPRESS): a multicentre, pragmatic, open-label, randomised controlled trial. *Intensive Care Med.* 2025;51(5):883-892.
7. Hernández G, Ospina-Tascón GA, Damiani LP, Estenssoro E, Dubin A, Hurtado J, et al. Effect of a Resuscitation Strategy Targeting Peripheral Perfusion Status vs Serum Lactate Levels on 28-Day Mortality Among Patients With Septic Shock: The ANDROMEDA-SHOCK Randomized Clinical Trial. *JAMA.* 2019;321(7):654-664.
8. The ANDROMEDA-SHOCK-2 Investigators for the ANDROMEDA Research Network, Spanish Society of Anesthesiology, Reanimation and Pain Therapy (SEDAR), and Latin American Intensive Care Network (LIVEN). Personalized Hemodynamic Resuscitation Targeting Capillary Refill Time in Early Septic Shock: The ANDROMEDA-SHOCK-2 Randomized Clinical Trial. *JAMA.* 2025;334(22):1988-1999.

PIC

Lemiale V, Resche-Rigon M, Zerbib Y, Mokart D, De Prost N, Wallet F, et al. Adjunctive corticosteroids in non-AIDS patients with severe *Pneumocystis jirovecii* pneumonia (PIC): a multicentre, double-blind, randomised controlled trial. Lancet Respir Med. 2025;13:800–808.

Introduction

Pneumocystis jirovecii pneumonia is an important cause of respiratory failure in patients with impaired cell-mediated immunity. The population at risk extends well beyond HIV infection to include haematological and solid malignancy, transplantation and treatment of inflammatory disease. These patients often have several overlapping causes of immunosuppression, and corticosteroids themselves may be part of the exposure that allowed the infection to develop.

Adjunctive corticosteroids improve outcomes in HIV-associated *Pneumocystis* pneumonia with substantial hypoxaemia. That evidence supplies a compelling rationale for considering the same treatment in other immunocompromised patients, but does not settle the question. Non-HIV disease may progress more rapidly, with greater pulmonary inflammation despite a lower organism burden. Underlying illness, concurrent infection and pre-existing steroid exposure can also alter both benefit and harm.¹

The proposed benefit is reduction of lung injury accompanying the host response and treatment of the organism. The competing concern is further suppression of antimicrobial immunity, alongside hyperglycaemia and other steroid-related effects. Observational studies have produced conflicting associations, partly because patients receiving adjunctive steroids may be more severely ill than those who do not. A randomised comparison was needed to separate treatment effects from treatment selection.

PIC provides the first substantial placebo-controlled trial addressing this question directly in HIV-negative patients. Its primary result narrowly missed the trial's significance threshold, while several secondary outcomes favoured treatment. Interpreting that pattern requires more than assigning the study a binary positive or negative label.

Synopsis

PIC was a multicentre, double-blind randomised trial conducted at 27 French hospitals. Recruitment extended from February 2017 to February 2024. The French Ministry of Health funded the study and had no role in its design, conduct, analysis or reporting. An independent data and safety monitoring board reviewed interim results. The principal report was published on 10 July 2025 and corrected on 1 August 2025.¹

Adults were eligible with acute hypoxaemic respiratory failure attributed to *Pneumocystis pneumonia*, microbiological documentation and fewer than seven days of anti-*Pneumocystis* treatment. The main report defined hypoxaemia as room-air PaO₂ below 60 mmHg or a need for more than 3 L/min oxygen to maintain saturation above 92%. Patients with HIV infection were excluded, as were those receiving more than 1 mg/kg/day corticosteroid for an underlying condition or with a contraindication to the trial treatment.

Diagnosis combined clinical and radiological probability with immunofluorescence or PCR on bronchoalveolar lavage fluid or induced sputum. PCR positivity alone was accepted only with a high clinical probability. This allowed inclusion without direct microscopic demonstration, while recognising that detection of the organism must be interpreted in the context of disease rather than colonisation.

Web-based randomisation allocated patients equally to intravenous methylprednisolone or saline placebo. Allocation used fixed-size blocks unknown to local investigators, with stratification by centre, preceding long-term corticosteroid treatment, malignancy versus another underlying condition, and oxygen requirement below or at least 6 L/min. Preparation was performed by a pharmacist or nurse outside patient care. Patients, investigators and treating staff were masked.

Methylprednisolone was prescribed at 30 mg twice daily for days one to five, 30 mg daily for days six to ten and 20 mg daily thereafter to day 21. Matching placebo followed the same schedule. Trial medication continued until day 21 or hospital discharge. Patients already taking corticosteroids below the exclusion threshold continued their baseline dose, so the intervention was additional treatment rather than necessarily steroids versus no steroid exposure.

The primary outcome was all-cause mortality at day 28. Secondary outcomes included mortality at day 90, intubation after randomisation, organ failure, infections, insulin requirements and length of stay. Analyses of intubation and infection accounted for death as a competing event. The initial sample calculation assumed mortality of 40%

with placebo and 20% with corticosteroids, an unusually large anticipated absolute reduction.

An additional interim analysis increased planned recruitment to 226. The final primary significance threshold was $P \leq 0.044$ under the sequential design. The report's intention-to-treat population excluded patients withdrawing consent or immediately found to have been enrolled in error. This is a modified allocation-based population rather than every randomised patient.

Of 466 screened patients, 226 were randomised: 114 placebo and 112 corticosteroid. Seven withdrew consent and one was enrolled by mistake, leaving 111 and 107 for analysis. Follow-up to day 90 was complete in those retained. Median age was 67 years, and 95% were in an ICU or intermediate-care unit at randomisation. Approximately 49% already received long-term corticosteroids, and 41% were invasively ventilated.

Pneumocystis was documented by PCR alone in 83%. Pulmonary coinfection was present in 17%. Antimicrobial treatment had started a median of three days before randomisation. Although the planned schedule lasted 21 days, typical trial treatment exposure was approximately thirteen days, reflecting discharge, death and other reasons for stopping. Eight placebo patients received additional corticosteroids during the first week, and two intervention patients received doses above those specified.

At day 28, 23/107 corticosteroid patients (21.5%) and 36/111 placebo patients (32.4%) had died. The reported absolute reduction, expressed as placebo minus corticosteroid mortality, was 10.9 percentage points (95% CI -0.9 to 22.5; $P=0.069$). This did not meet the primary threshold. A post hoc adjustment for stratification variables likewise did not establish significance, with an odds ratio of 0.58 (95% CI 0.30–1.10).

At day 90, deaths numbered thirty (28.0%) versus 48 (43.2%). The mortality hazard ratio was 0.59 (95% CI 0.37–0.93; $P=0.022$). Among the 129 patients not intubated at baseline, six of sixty corticosteroid recipients and eighteen of 69 placebo recipients subsequently required intubation; the cause-specific hazard ratio was 0.36 (95% CI 0.14–0.90; $P=0.020$). Median ventilator-free days were 25 versus nineteen.

Secondary infections occurred in 25/107 corticosteroid patients (23.4%) and 38/111 placebo patients (34.2%). Increased insulin needs occurred in 33/107 (30.8%) and 25/111 (22.5%), respectively. Neither comparison established a clear difference under the reported tests. No clear advantage was demonstrated for other organ failure or length-of-stay outcomes.^{1,2}

Critique

PIC is a valuable trial in a difficult-to-recruit population. Masking, placebo control, central allocation and complete mortality follow-up among retained patients strengthen its internal validity. It provides much more direct evidence than extrapolation from HIV trials or comparisons of treated and untreated observational cohorts. Its main limitation is the precision of the clinical effect, not an absence of any favourable signal.

The day-28 result should not be described as proof that corticosteroids do not work. The point estimate represents a substantial mortality reduction, and the interval extends from a small increase to a very large reduction. The data are therefore compatible with clinically important benefit. Equally, a P value of 0.069 cannot be turned into a positive primary result by describing it as almost significant. The trial specified a threshold of 0.044, and that criterion was not met.

The sample calculation anticipated halving mortality from 40% to 20%. Actual placebo mortality was lower, and the observed relative reduction was less extreme. A trial sized for a twenty-percentage-point absolute effect will have limited power for smaller, still important benefits. This matters when interpreting the negative hypothesis test, but does not licence certainty that the favourable point estimate is the true effect. The wide interval is the appropriate expression of what remains unresolved.

Day-90 mortality adds meaningful information. Follow-up was complete, deaths are objective events, and the direction is consistent with the primary result. The later difference was also larger in absolute terms. However, day 90 was a secondary endpoint among numerous comparisons, without a reported multiplicity framework making it an independent confirmatory success after the primary test failed. It strengthens the possibility of benefit without replacing the prespecified primary conclusion.

The two mortality outcomes are correlated rather than independent replications. Most deaths contributing to day 28 also contribute to day 90. Likewise, intubation, ventilator-free days and survival share aspects of the same clinical trajectory. Several favourable results make a coherent account more plausible, but counting each as a separate confirmation exaggerates the amount of information in 218 patients.

The intubation result is clinically attractive and reasonably analysed using a baseline-defined subset. Unlike restricting analysis to a subgroup selected by post-randomisation treatment response, excluding those already intubated does not itself condition on an effect of the intervention. Nevertheless, only 24 new intubations occurred, and the estimate remains imprecise. Treatment-limitation decisions were

present in some patients at baseline, which can affect the relationship between respiratory deterioration and actual intubation.

Competing-risk methods help distinguish death before intubation from freedom from respiratory failure, but the cause-specific hazard ratio is not a simple ratio of cumulative risks. The six of sixty versus eighteen of 69 counts make the observed comparison clearer. It would be misleading to use the full group denominators for new intubation, as one paragraph of the main report does; the table footnote and subsequent text identify the correct at-risk population.

Previous steroid exposure changes the treatment contrast. Approximately half the cohort continued corticosteroids for their underlying disease. In these patients, PIC tested the effect of adding a defined course to existing treatment. It did not test withdrawal, continuation or initiation from zero exposure. The dose ceiling also excluded many patients already receiving intensive immunosuppression. Results cannot be assumed to apply to escalating treatment in those taking more than 1 mg/kg/day.

Timing is another important distinction from the simple phrase “early corticosteroids”. Patients entered after a median of three days of antimicrobial therapy and could have received treatment for nearly a week. Waiting for microbiological evidence improved diagnostic confidence but may have missed part of the initial inflammatory response. Conversely, empirical treatment before confirmation may expose patients with an alternative diagnosis to unnecessary steroids. PIC does not determine the best balance between immediate empirical treatment and diagnostic certainty.

Diagnostic inclusion was pragmatic but imperfect. PCR alone established microbiological positivity in most patients, and clinicians required compatible imaging and high pretest probability to infer pneumonia. This is a reasonable approach in a high-risk cohort, but residual misclassification is possible because PCR can detect colonisation. Concurrent bacterial, viral or fungal infection could also influence the clinical course and the response to steroids. Randomisation balances such uncertainty on average, without identifying which biological phenotype benefits.

The population was heterogeneous. Malignancy, transplantation and inflammatory disease differ in immune defects, background drugs and competing mortality risks. A post hoc analysis appeared more favourable in patients without malignancy, but the interaction test did not establish that the treatment effect differed by malignancy status. A confidence interval excluding one in one subgroup and crossing one in another is not itself evidence of different effects. PIC does not provide a validated rule to offer treatment only to one underlying-disease category.²

Safety findings are encouraging but limited. There was no observed excess of secondary infection, and infection counts numerically favoured corticosteroids. The trial was nevertheless too small to exclude uncommon opportunistic complications or to establish broad safety across all forms of immunosuppression. Insulin requirements were numerically higher with treatment, consistent with a recognised physiological effect, but were also imprecisely estimated. “No significant difference” should not be translated into absence of harm.

Infections were defined clinically by a need for antibiotics for a new infection. Blinding reduces the likelihood that treatment knowledge changed that judgement, but diagnostic uncertainty remains. Death and discharge also affect time at risk and observation. The report presents both counts and cumulative-incidence-derived comparisons; these are different estimands and do not always yield numerically identical differences. The safest summary states the counts and acknowledges uncertainty rather than reconstructing an effect from incompatible table entries.

Treatment duration should also be kept distinct from the prescribed schedule. Many patients did not receive all 21 days. Shorter exposure may be appropriate after discharge or unavoidable after death, and should not automatically be labelled poor adherence. The trial estimates the effect of offering the schedule under those stopping rules. It does not establish that thirteen days is an optimal course, nor compare intravenous treatment with an equivalent oral regimen.

Seven years of recruitment across 27 centres illustrates both the rarity of eligible disease and the effort needed to obtain randomised evidence. It also spans changes in immunosuppressive treatments, respiratory support and the COVID-19 period. Randomisation remains the main protection against imbalance over that time, but modest numbers limit exploration of changing case mix. The results are most applicable to hospitalised patients with clinically important hypoxaemia resembling those enrolled.

The accompanying editorial favours considering corticosteroids and proposes pathogen burden as a future means of selecting patients. That is a reasonable hypothesis to investigate, not a selection strategy tested by PIC. The trial did not establish a quantitative organism threshold or demonstrate that such a threshold predicts benefit. Clinical interpretation should separate the randomised findings from the editorial's mechanistic model.³

PIC consequently shifts the evidence towards a possible benefit while leaving the effect size uncertain. For an individual patient with severe hypoxaemic non-HIV Pneumocystis pneumonia, the favourable direction, secondary outcomes and preceding evidence may

support consideration of adjunctive treatment. That decision should retain awareness of baseline steroid exposure, coinfection and contraindications. It should not be presented as a universally proven mortality intervention or extrapolated to mild non-hypoxaemic infection.

Body of Evidence

The HIV evidence began with trials showing substantial reductions in respiratory failure and death. In the multicentre study by Bozzette and colleagues, 333 patients with AIDS and suspected *Pneumocystis pneumonia* were randomised to adjunctive corticosteroids or standard treatment alone. Among those with confirmed or presumed disease, day-31 respiratory failure and mortality were lower with steroids. Benefit was not demonstrated in mild disease, and localised herpetic lesions were more frequent.⁴

Gagnon and colleagues conducted a smaller double-blind trial in severe AIDS-associated disease, enrolling patients within 72 hours of starting antimicrobials. Nine of twelve methylprednisolone recipients survived to discharge compared with two of eleven placebo recipients. The trial stopped early because of the large observed difference. Its size and early stopping limit precision, but its direction complemented the larger HIV trial rather than standing alone.⁵

A Cochrane review incorporating six adult trials in its meta-analysis found lower mortality at one month and at three to four months, with risk ratios of 0.56 (95% CI 0.32–0.98) and 0.59 (95% CI 0.41–0.85), respectively. The three largest trials also supported less mechanical ventilation. These results provide a sound basis for the HIV indication with substantial hypoxaemia. They are prior evidence for PIC's biological rationale, while the host population and background care remain different.⁶

Those differences have empirical support. Limper and colleagues studied 75 patients with *Pneumocystis pneumonia* and found greater organism numbers but fewer lavage neutrophils in AIDS than in other immunocompromised conditions. Greater neutrophilic inflammation correlated with worse oxygenation and survival. This observational work supports an important role for host response, but cannot show whether suppressing that response improves outcomes in every non-HIV condition.⁷

Before PIC, a meta-analysis of sixteen retrospective non-HIV cohorts, comprising 2,518 cases, found apparently conflicting results. Adjunctive steroids were associated with higher mortality in the overall population, but lower mortality in the subgroup with respiratory failure. Differences in severity, timing and treatment selection make both associations vulnerable to confounding. Clinicians may prescribe steroids because a

patient is deteriorating, while restriction to more severely hypoxaemic patients can change the apparent balance.⁸

PIC materially improves on that evidence by randomising treatment within a defined hypoxaemic population and masking both care and outcome assessment. Its direction is consistent with the favourable severe-disease observational subgroup and the HIV trials, but it does not reproduce a definitive primary mortality result. The most defensible synthesis is therefore a plausible clinical benefit with substantial remaining uncertainty, rather than either automatic extrapolation from HIV or dismissal of steroids because one threshold was missed.

Future work should clarify which patients benefit, the importance of preceding steroid exposure and the timing and duration of additional treatment. Quantitative organism burden, immune phenotype and coinfection may help define more coherent populations, but need prospective validation. Until then, the PIC regimen is evidence about a specific adjunctive strategy in selected hypoxaemic adults, not a general licence for higher-dose immunosuppression.

Should adjunctive corticosteroids be considered for severe non-HIV *Pneumocystis pneumonia*?

PIC supports consideration in patients resembling those enrolled, with a favourable but inconclusive day-28 mortality estimate and supportive secondary outcomes. It did not establish a definitive primary mortality benefit or a validated subgroup-selection rule.

References

1. Lemiale V, Resche-Rigon M, Zerbib Y, Mokart D, De Prost N, Wallet F, et al. Adjunctive corticosteroids in non-AIDS patients with severe *Pneumocystis jirovecii* pneumonia (PIC): a multicentre, double-blind, randomised controlled trial. *Lancet Respir Med*. 2025;13(9):800-808.
2. Lemiale V, Resche-Rigon M, Zerbib Y, Mokart D, De Prost N, Wallet F, et al. PIC supplementary appendix: eligibility and results, appendices 3–8, printed pages 50–56. *Lancet Respir Med*. 2025. doi:10.1016/S2213-2600(25)00125-0.
3. Shapiro L, Henao-Martínez AF. Adjunct corticosteroids for non-HIV infected patients with severe *Pneumocystis jirovecii* pneumonia. *Lancet Respir Med*. 2025;13:771–773.

4. Bozzette SA, Sattler FR, Chiu J, Wu AW, Gluckstein D, Kemper C, et al. A controlled trial of early adjunctive treatment with corticosteroids for *Pneumocystis carinii* pneumonia in the acquired immunodeficiency syndrome. California Collaborative Treatment Group. *N Engl J Med*. 1990;323(21):1451-1457.
5. Gagnon S, Boota AM, Fischl MA, Baier H, Kirksey OW, La Voie L. Corticosteroids as adjunctive therapy for severe *Pneumocystis carinii* pneumonia in the acquired immunodeficiency syndrome. A double-blind, placebo-controlled trial. *N Engl J Med*. 1990;323(21):1444-1450.
6. Ewald H, Raatz H, Boscacci R, Furrer H, Bucher HC, Briel M. Adjunctive corticosteroids for *Pneumocystis jiroveci* pneumonia in patients with HIV infection. *Cochrane Database Syst Rev*. 2015;(4):CD006150.
7. Limper AH, Offord KP, Smith TF, Martin WJ 2nd. *Pneumocystis carinii* pneumonia. Differences in lung parasite number and inflammation in patients with and without AIDS. *Am Rev Respir Dis*. 1989;140(5):1204-1209.
8. Ding L, Huang H, Wang H, He H. Adjunctive corticosteroids may be associated with better outcome for non-HIV *Pneumocystis* pneumonia with respiratory failure: a systemic review and meta-analysis of observational studies. *Ann Intensive Care*. 2020;10(1):34.

SuDDICU

Cuthbertson BH, Billot L, Campbell MK, Daneman N, Davis JS, Delaney A, et al. Selective Decontamination of the Digestive Tract during Ventilation in the ICU. N Engl J Med. 2026;394(15):1491-1502. First published online 29 October 2025.

Introduction

Selective decontamination of the digestive tract has remained controversial despite decades of investigation. The intervention aims to suppress potentially pathogenic organisms in the mouth and gastrointestinal tract before they cause infection in patients receiving mechanical ventilation. Topical antimicrobials are combined with a short systemic antibiotic course, targeting a reservoir from which lower respiratory and bloodstream infection may arise.

The attraction is preventive rather than therapeutic: avoiding infection might reduce organ dysfunction, antibiotic treatment and death. The concern is that widespread antimicrobial exposure could alter resistance within an ICU, including among patients who never receive the intervention. Individual benefit and ecological harm are therefore different questions. A trial that measures only infections in treated patients cannot fully answer whether adopting the strategy is beneficial for an entire unit.

SuDDICU was designed around both questions. Its Australian component had already been published in 2022. The full international report added Canadian participants using the same intervention and trial framework, producing a larger estimate of patient outcomes alongside a contemporaneous assessment of ICU microbial ecology. The Australian results are part of this combined dataset, not an independent replication of it.^{1,2}

Synopsis

SuDDICU was an international, open-label, cluster-randomised crossover trial in twenty-six general medical and surgical ICUs: nineteen in Australia and seven in Canada. Each ICU adopted SDD or standard care for twelve months, followed by a three-month gap and twelve months using the other strategy. Randomisation was stratified by ICU bed numbers. The crossover occurred at unit level; individual patients were not deliberately switched between treatments as part of a paired comparison.¹

The programme was publicly funded in Australia and Canada. Trial-specific oral and gastric preparations were manufactured under contract. Australian participation used a waiver of individual consent. In Canada, deferred informed consent was required for SDD recipients, while consent was waived for standard-care patients and the ecological assessment. That difference in enrolment procedures became an important issue when interpreting the combined results.

Patients qualified if invasive ventilation through an endotracheal tube was expected to continue beyond the end of the following calendar day, whether ventilation began on ICU admission or subsequently. Patients who remained ventilated beyond an initially shorter prediction could also become eligible. This selected a population at sustained risk of ICU-acquired infection rather than every briefly intubated patient. Standard care followed local practice, including institutional antibiotic-stewardship policies.

SDD comprised 0.5 g of oral paste containing colistin 10 mg, tobramycin 10 mg and nystatin 125,000 IU every six hours, applied to the oropharynx and buccal mucosa. A 10 mL gastric or postpyloric suspension supplied colistin 100 mg, tobramycin 80 mg and nystatin two million IU at the same interval. Topical treatment continued during mechanical ventilation or until day ninety. A four-day intravenous course, usually a third-generation cephalosporin or ciprofloxacin, was added unless the patient was already receiving an antibiotic with suitable gram-negative activity.¹

The primary outcome was death before discharge from the index hospital admission, censored at ninety days. It was not all-cause mortality wherever the patient happened to be at day ninety. Four secondary clinical outcomes were ICU mortality and days alive and free of ventilation, ICU admission and hospitalisation. Microbiological outcomes included new bloodstream infections, new positive cultures of prespecified resistant organisms and new *Clostridioides difficile* infection. Intravenous antibiotic consumption was measured over twenty-eight days.

The ecological assessment enrolled other ICU patients during one week per month in five three-month windows: before the trial, late in the first treatment period, during the gap, late in the second period and after the trial. Patients already included in the clinical comparison were excluded from this additional cohort. The assessment therefore examined possible effects extending beyond SDD recipients, using cultures obtained in routine care rather than mandatory cultures from every patient.^{1,3}

The main report's sample-size calculation anticipated 9,000 patients from thirty units, providing 80% power for a 3.5-percentage-point mortality reduction from 29%. The primary hierarchical logistic model incorporated both cluster and cluster-period effects,

with a correction for limited degrees of freedom. Analyses followed the assigned unit and period regardless of adherence. Four clinical secondary outcomes used a step-down Holm–Bonferroni procedure. Microbiological, subgroup and ecological intervals were not adjusted for multiple comparisons.

For ecology, the acceptable excess was two percentage points. Changes over time were compared between the two unit sequences in two specified comparisons, incorporating the intervention periods and adjacent gaps. Noninferiority required the upper confidence limit for the excess with SDD to remain below that margin. This was a separate noninferiority question, not a reinterpretation of the mortality superiority analysis.¹

Overall, 20,000 patients contributed: 9,289 to the clinical trial and 10,711 to ecology. The clinical population comprised 5,982 Australian and 3,307 Canadian patients, with 4,223 enrolled during SDD periods and 5,066 during standard-care periods. Mean age was approximately 59 years and 36% were female. Illness severity was assessed with APACHE scores in Australia and the Multiple Organ Dysfunction Score in Canada. Primary-outcome data were available for 9,280 participants, or 99.9%.

At least one daily dose of both topical components was delivered on 86.2% of eligible ventilation days. This was a daily exposure measure, rather than confirmation that all four scheduled administrations occurred. Suitable intravenous antibiotics were reported during the first four days in approximately 84% of SDD recipients and 59% of controls. Thus the main treatment contrast was substantial topical exposure plus an increase in early systemic coverage, against a background in which many controls already received active antibiotics.³

There were 1,175 deaths among 4,215 SDD patients with known outcomes, 27.9%, and 1,494 among 5,065 controls, 29.5%. The primary model gave an odds ratio of 0.94 (95% CI 0.84–1.05; $P=0.27$). The adjusted absolute difference was –1.30 percentage points (95% CI –3.60 to 1.00). Further adjustment for baseline clinical characteristics produced an odds ratio of 0.93 (95% CI 0.84–1.04). The primary comparison therefore did not establish a mortality reduction.¹

ICU mortality was 22.0% versus 24.2%. Mean days alive and free of ventilation were 61.0 versus 58.9, with an adjusted difference of 1.75 days (95% CI –0.20 to 3.70). The adjusted difference in ICU-free days was 1.96 (95% CI 0.03–3.90), and in hospital-free days 0.93 (95% CI –0.90 to 2.70). None of the four clinical secondary comparisons was significant after the specified multiplicity procedure, including ICU-free days despite its nominal interval excluding zero.

New bloodstream infections occurred in 208 patients, 4.9%, versus 347, 6.8%, adjusted difference –1.30 percentage points (95% CI –2.55 to –0.05). New resistant-organism cultures were reported in 709, 16.8%, versus 1,359, 26.8%, adjusted difference –9.60 percentage points (95% CI –12.40 to –6.80). New *C. difficile* infection occurred in 0.7% versus 1.2%, with an interval spanning no difference. Overall intravenous antibiotic consumption did not show a clear difference.^{1,3}

The ecological analysis met its noninferiority criterion for bloodstream infection and *C. difficile* infection. It did not do so for new resistant-organism cultures. The estimated differences for resistance were –2.05 percentage points (95% CI –7.42 to 3.31) in the first comparison and –0.33 (95% CI –5.48 to 4.82) in the second. Both upper limits exceeded the two-percentage-point margin. Post hoc alternative comparisons did not resolve that uncertainty.¹

Twelve adverse drug reactions, 0.3%, were reported with SDD and none with standard care; none was classified as serious. Serious adverse events were reported in 47 patients, 1.1%, versus 59, 1.2%. No clear heterogeneity appeared across the prespecified age, sex, illness-severity or trauma subgroups. A post hoc acute-brain-injury subgroup had a favourable mortality estimate, but remained an exploratory finding.¹

Critique

SuDDICU is unusually informative because it recognises that an antimicrobial prevention policy has consequences beyond the treated individual. Unit-level randomisation fits how SDD would be implemented, reduces contamination between intervention and control patients within a period, and permits examination of wider ecology. The clinical sample is large, outcome ascertainment nearly complete, and hospital mortality less vulnerable to diagnostic judgement than ventilator-associated pneumonia.

Nevertheless, the neutral primary comparison should be expressed as uncertainty rather than proof of no benefit. The absolute confidence interval includes a mortality reduction of 3.6 percentage points, which would be clinically substantial, as well as a one-point increase. The point estimate favours SDD, but the trial did not distinguish that estimate reliably from no effect. A large participant count does not guarantee precision when the unit of randomisation is the ICU and outcomes within units are correlated.

Only twenty-six clusters participated, fewer than the thirty in the main report's planning calculation. The observed intracluster correlation was also greater than initially assumed. These features reduce effective information relative to an individually randomised study of the same size. At the same time, control mortality closely matched the anticipated

29%, so unexpectedly low baseline risk is not the explanation for the result. The detectable effect and the uncertainty remaining should be judged from the model and interval, not the headline number of patients.

The crossover design lets each ICU act as its own control, helping account for stable differences in case mix and practice. It also introduces dependence on time. Recruitment spanned several years, and treatment, admission patterns and infection epidemiology could change between periods. A three-month gap reduces immediate overlap but cannot guarantee that ecological effects disappear. Persistent changes in colonisation or resistance could carry into the later period and alter the apparent contrast.

Unequal recruitment is another important limitation. The SDD group was smaller, particularly in Canada, where the intervention required deferred consent but control care did not. The flow diagrams show many eligible patients not enrolled and many exclusions with undocumented reasons. This creates a route for selection after the unit's assignment is known. Similar measured baseline averages do not exclude differences in prognosis, willingness to participate or timing of enrolment that were not captured.

The main model and sensitivity analyses help, but adjustment cannot recreate random selection of individuals within an assigned cluster. One post hoc analysis excluding patients enrolled within an hour of ICU admission produced a more favourable mortality interval. That result is useful for exploring how enrolment timing might matter, but the exclusion changes the population and was not the primary comparison. It should not be selected as the definitive estimate because it crosses a statistical threshold.³

The mortality endpoint is robust within its stated boundary. Death after discharge was outside the index-admission primary outcome. This matters if an intervention changes discharge timing or transfers risk beyond hospital, although the trial did not demonstrate a major hospital-free-day advantage. The chapter's mortality language therefore remains hospital-specific. A claim of improved ninety-day survival in all settings would be broader than the measurement.

Adherence was encouraging but is easily overstated. Receiving at least one dose of each topical preparation on a day does not mean completing every six-hourly dose. Some patients had delayed initiation or omitted treatment. These deviations can reduce the achieved contrast, but the intention-to-treat effect is also relevant to real implementation. A hypothetical benefit under perfect administration cannot be inferred from the neutral policy comparison.

The systemic antibiotic component complicates mechanistic interpretation. Many controls already received suitable gram-negative coverage, while more SDD patients received it early. The trial tests this combined strategy against contemporary usual care, not topical antibiotics in isolation. Similar total intravenous defined daily doses could reflect differences in timing, indication or drug selection. It also excludes the topical antimicrobial burden and therefore does not establish identical total antibiotic exposure or identical selection pressure.

The microbiological findings are consistent with suppression of detectable organisms and fewer bloodstream infections, but several cautions remain. Routine clinical cultures are influenced by whether staff suspect infection and which specimens they obtain. Blood cultures were collected in 59.7% of SDD patients and 66.4% of controls. Fewer cultures may follow genuinely fewer infections, altered clinical suspicion, or both. In an open-label trial, these pathways cannot be separated simply by comparing positive results.³

A positive resistant-organism culture is also not synonymous with a clinically important resistant infection. Conversely, preventing colonisation can still be valuable by reducing a reservoir for subsequent infection. The endpoint should retain its actual name rather than being translated into either proven prevention of resistant disease or an irrelevant laboratory change. The intervals were not multiplicity-adjusted, so these findings provide supportive evidence without replacing the primary clinical outcome.

The apparent contrast between fewer resistant cultures in SDD recipients and inconclusive ecological safety is not a contradiction. The analyses concern different patients and different effects. Suppression during treatment can reduce carriage detected in recipients while leaving uncertain the longer-term selection and transmission of resistant organisms across the unit. The ecological cohort was specifically intended to address that wider question.

Failure of ecological noninferiority does not show that SDD increased resistance. Both point estimates favoured SDD, but the intervals also allowed increases exceeding the accepted margin. The correct conclusion is that the study could not exclude that degree of ecological harm. Short sampled windows, few clusters and reliance on routine cultures limit reassurance about sustained adoption over years, particularly in a setting with a different baseline resistance pattern.

The brain-injury finding is attractive because these patients may remain ventilated while their immediate prognosis is determined by neurological disease rather than infection. However, the subgroup was post hoc, the confidence intervals were not

multiplicity-adjusted, and the combined analysis includes the Australian patients who generated the earlier hypothesis. It is not an independent confirmation of a brain-specific benefit. Differences between subgroup P values would not by themselves establish an interaction.^{1,4}

The four clinical secondary outcomes also illustrate why prespecified analysis matters. A nominal ICU-free-day interval just excluded zero, but none remained positive after the stated multiplicity procedure. Promoting this isolated result would overstate recovery benefit. Equally, the consistent favourable direction of several clinical estimates can inform an overall assessment without being presented as a series of confirmed effects.

Generalisation depends on local ecology and care. Australia and Canada provided broad medical and surgical case mixes but relatively low rates of resistant infections. An ICU facing endemic resistance to the topical agents may have a different balance of effect. Commercial-grade preparations, nursing workload, six-hourly delivery and surveillance also matter. SuDDICU supplies evidence for a defined policy under those conditions, not a universal assurance that any decontamination mixture will work safely everywhere.

Body of Evidence

In a 2003 single-centre randomised trial of 934 patients, de Jonge and colleagues found lower hospital mortality with SDD, 24% versus 31%, and less acquisition of resistant gram-negative organisms. The setting had very low methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant enterococcal prevalence. The result supported efficacy under favourable ecological conditions but left uncertainty about transportability to other units and eras.⁵

A Dutch cluster crossover trial in thirteen ICUs subsequently compared SDD, selective oropharyngeal decontamination and standard care in 5,939 patients. Crude twenty-eight-day mortality differed little, but covariate-adjusted odds ratios favoured both decontamination strategies; for SDD the estimate was 0.83 (95% CI 0.72–0.97). This demonstrates why crude event rates and model-adjusted effects should not be mixed, and why baseline imbalance and cluster structure are important to interpretation.⁶

R-GNOSIS extended investigation to thirteen European ICUs with moderate or high resistance. It included 8,665 patients across baseline and several intervention periods and did not establish reductions in its primary resistant gram-negative bloodstream-infection outcome. Its SDD regimen did not include the routine short intravenous component used in SuDDICU. Differences in background resistance,

systemic treatment and study comparisons limit direct equivalence between the programmes.⁷

The Australian SuDDICU report included 5,982 patients and found hospital mortality of 27.0% versus 29.1%, odds ratio 0.91 (95% CI 0.82–1.02). It also showed fewer new positive blood and resistant-organism cultures but failed to establish ecological noninferiority for resistance. The full international report adds information to that estimate; counting both publications as separate trials would duplicate those patients.²

A 2022 systematic review included thirty-two trials with 24,389 participants, of which thirty contributed mortality data. Its Bayesian analysis estimated a mortality risk ratio of 0.91 (95% credible interval 0.82–0.99), with a 99.3% posterior probability of lower mortality under the model. Evidence on resistance was much less certain. Trials using an intravenous component appeared more favourable, but that between-trial comparison does not isolate the causal contribution of the intravenous drug.⁸

The favourable pooled estimate and the full trial's inconclusive mortality result can coexist. They combine different amounts of information, populations and assumptions, while much of the SuDDICU evidence overlaps. The contemporary trial narrows the question to a reproducible unit policy and exposes the remaining ecological uncertainty. It supports continued consideration of a possible clinical benefit, while withholding the definitive patient and unit-level reassurance needed for an uncomplicated universal recommendation.

Does the full SuDDICU trial establish that routine SDD improves survival without increasing ICU resistance?

No. It leaves a clinically important hospital-mortality benefit possible, while failing to establish ecological noninferiority for resistant-organism cultures.

References

1. Cuthbertson BH, Billot L, Campbell MK, Daneman N, Davis JS, Delaney A, et al. Selective Decontamination of the Digestive Tract during Ventilation in the ICU. *N Engl J Med*. 2026;394(15):1491-1502. First published online in 2025.
2. Myburgh JA, Seppelt IM, Goodman F, Billot L, Correa M, Davis JS, et al. Effect of Selective Decontamination of the Digestive Tract on Hospital Mortality in Critically Ill Patients Receiving Mechanical Ventilation: A Randomized Clinical Trial. *JAMA*. 2022;328(19):1911-1921.

3. Cuthbertson BH, Billot L, Campbell MK, Daneman N, Davis JS, Delaney A, et al. Supplementary methods and results to Selective Decontamination of the Digestive Tract during Ventilation in the ICU. *N Engl J Med*. Published online 2025.
4. Young PJ, Devaux A, Li Q, Billot L, Davis JS, Delaney A, et al. Selective digestive tract decontamination in critically ill adults with acute brain injuries: a post hoc analysis of a randomized clinical trial. *Intensive Care Med*. 2024;50(1):56-67.
5. de Jonge E, Schultz MJ, Spanjaard L, Bossuyt PM, Vroom MB, Dankert J, et al. Effects of selective decontamination of digestive tract on mortality and acquisition of resistant bacteria in intensive care: a randomised controlled trial. *Lancet*. 2003;362(9389):1011-1016.
6. de Smet AM, Kluytmans JA, Cooper BS, Mascini EM, Benus RF, van der Werf TS, et al. Decontamination of the digestive tract and oropharynx in ICU patients. *N Engl J Med*. 2009;360(1):20-31.
7. Wittekamp BH, Plantinga NL, Cooper BS, Lopez-Contreras J, Coll P, Mancebo J, et al. Decontamination Strategies and Bloodstream Infections With Antibiotic-Resistant Microorganisms in Ventilated Patients: A Randomized Clinical Trial. *JAMA*. 2018;320(20):2087-2098.
8. Hammond NE, Myburgh J, Seppelt I, Garside T, Vlok R, Mahendran S, et al. Association Between Selective Decontamination of the Digestive Tract and In-Hospital Mortality in Intensive Care Unit Patients Receiving Mechanical Ventilation: A Systematic Review and Meta-analysis. *JAMA*. 2022;328(19):1922-1934.

ImmunoSep

Giamarellos-Bourboulis EJ, Kotsaki A, Kotsamidi I, Efthymiou A, Koutsoukou V, Ehler J, et al. Precision Immunotherapy to Improve Sepsis Outcomes: The ImmunoSep Randomized Clinical Trial. JAMA. 2026;335(9):775-786. First published online 8 December 2025.

Introduction

Sepsis can involve excessive inflammatory injury and impaired antimicrobial immunity at the same time. This creates a fundamental problem for treatment: suppressing a damaging host response may help one patient while worsening another patient's inability to control infection. A clinical diagnosis of sepsis identifies organ dysfunction caused by infection, but does not necessarily identify the immune process that should be targeted.

Many earlier immunomodulatory trials treated broad sepsis populations with one intervention. Their disappointing results encouraged a different approach: select patients using biological measurements, then match treatment to the abnormality. This requires more than showing that a biomarker predicts death. The marker must identify a clinically useful treatment decision, and the testing process must be fast enough to act while that decision remains relevant.

ImmunoSep tested a practical version of this idea. Ferritin and monocyte HLA-DR expression defined two immune patterns, linked to either interleukin-1 blockade with anakinra or immune stimulation with interferon gamma. The randomised comparison was a biomarker-guided strategy against placebo, with early organ dysfunction as the principal outcome. It was an important test of feasibility and biological effect, while remaining a phase 2b study rather than a definitive survival trial.¹

Synopsis

ImmunoSep was a multicentre, double-blind, double-dummy, placebo-controlled trial at thirty-three sites in Greece, Germany, Italy, the Netherlands, Switzerland and Romania. Recruitment began in August 2021, and final follow-up ended in April 2024. The European Union Horizon 2020 programme funded the study; the Hellenic Institute for the Study of Sepsis sponsored and conducted it, including the analysis and interpretation. Patients or appropriate relatives provided written consent.¹

Eligible adults had Sepsis-3 sepsis arising from community-acquired, hospital-acquired or ventilator-associated pneumonia, or primary bacteraemia. Treatment had to begin within seventy-two hours of sepsis onset. Major exclusions included abdominal and urinary sources, neutropenia, advanced malignancy, specified pre-existing immune disorders and substantial prolonged corticosteroid or recent biological immunosuppressive treatment. Active COVID-19 pneumonia could qualify only through the macrophage-activation-like pathway. This was a deliberately restricted infection population.^{1,2}

Blood was sent within six hours to a designated national laboratory. Ferritin above 4,420 ng/mL defined macrophage activation-like syndrome. Ferritin at or below that threshold together with fewer than 5,000 HLA-DR molecules per CD45/CD14 monocyte defined sepsis-induced immunoparalysis. Patients with lower ferritin and HLA-DR at least 5,000 were unclassified and excluded. High ferritin took precedence when both abnormalities were present; normal HLA-DR was not required for the macrophage-activation-like category.²

Computer-generated randomisation was performed separately by site, allocating patients to the active precision strategy or placebo in addition to standard care. Patients in the active group with high ferritin received anakinra 200 mg intravenously every eight hours. Those with immunoparalysis received recombinant human interferon gamma 100 micrograms subcutaneously every forty-eight hours. Treatment lasted up to fifteen days. The intravenous dose was halved when calculated creatinine clearance was below 30 mL/min.

All participants received both intravenous and subcutaneous injections. Active anakinra was paired with subcutaneous placebo, and active interferon with intravenous placebo; controls received both placebos. Unblinded pharmacists prepared identically labelled treatments, while patients, clinicians, nurses and investigators remained masked. This arrangement protected the comparison from knowledge of either the assigned strategy or the active drug being administered.

The primary endpoint was achieving a decrease of at least 1.4 points between baseline SOFA and the mean daily SOFA score over days two to nine. It was a binary responder outcome based on an average trajectory, not a 1.4-point difference between group means and not simply the SOFA score on day nine. SOFA ranges from zero to twenty-four across six organ systems. The threshold was motivated by earlier associations between changes in organ dysfunction and mortality.¹

Deaths were handled by assigning SOFA twenty-four from the day of death through day fifteen. After early discharge, the last observed score was carried forward. These rules

prevented death from appearing as disappearance of organ dysfunction, but also made the endpoint a mixture of measured physiology, survival and imputed values. Secondary outcomes included the response within each immune subgroup, twenty-eight- and ninety-day mortality, the same SOFA response through day fifteen, reversal of immune dysfunction and infection status at day fifteen.

Immune reversal required sustained ferritin reduction of at least 15% in the high-ferritin group, or HLA-DR above 8,000 molecules per monocyte in the immunoparalysis group, confirmed on serial samples. Infection was classified by masked site investigators as resolved, intermediate, failure or superinfection. The sample-size calculation assumed primary response in 40% versus 20%, requiring 117 per group for 90% power and a target of 280 after allowing for attrition.^{1,2}

The primary comparison used Fisher's exact test, with an odds ratio and confidence interval. A supplementary regression adjusted for baseline characteristics and site-related information. Secondary analyses used the same conventional significance threshold without a stated family-wise multiplicity correction. Survival and time to immune reversal were also examined using time-to-event methods. Additional severity subgroups were explored using cut points selected from the observed data.

Of 672 patients screened, 355 had an unclassified immune state and were excluded. There were 281 randomised participants, but five withdrew consent and requested removal of all data, leaving 131 in the precision group and 145 controls. Mean age was 70 years, one-third were female and almost all were White. Approximately 89% received mechanical ventilation and 17% continuous renal replacement therapy. Ventilator-associated pneumonia was the most common source.¹

Forty-eight patients had macrophage activation-like syndrome, twenty-five assigned anakinra and twenty-three placebo. The remaining 228 had immunoparalysis, with 106 assigned interferon and 122 placebo. Median baseline SOFA was ten versus nine in the overall active and control groups. Median time from sepsis onset to study treatment was thirty-six versus forty-seven hours. The tested strategy was therefore generally initiated after the first day of illness rather than at its earliest recognition.

The primary response occurred in 46/131 patients, 35.1%, with precision treatment and 26/145, 17.9%, with placebo: absolute difference 17.2 percentage points (95% CI 6.8–27.2), odds ratio 2.48 (95% CI 1.42–4.32; $P=0.002$). The adjusted odds ratio was 2.49 (95% CI 1.42–4.36). Excluding patients discharged before day nine gave a consistent sensitivity result. By that day, twenty-one active-group and twenty-nine control patients had died, while ten and thirteen respectively had carried-forward discharge values.^{1,2}

Within the high-ferritin subgroup, twelve of twenty-five anakinra recipients, 48.0%, responded, compared with four of twenty-three controls, 17.4%. In the immunoparalysis subgroup, response occurred in 34/106 interferon recipients, 32.1%, and 22/122 controls, 18.0%. The corresponding intervals were much wider than for the combined strategy, particularly in the small anakinra subgroup. At day fifteen, the overall response remained more frequent, 52/131, 39.7%, versus 34/145, 23.4%.^{1,2}

Twenty-eight-day mortality was 57/131, 43.5%, versus 72/145, 49.7%, odds ratio 0.78 (95% CI 0.49–1.26). Ninety-day mortality was 90/131, 68.7%, versus 98/145, 67.6%, odds ratio 1.05 (95% CI 0.63–1.75). Neither comparison established a difference. The numerical early mortality advantage therefore did not translate into a sustained separation at ninety days.¹

Only 125 participants contributed to the serial immune-reversal comparison. Reversal occurred in 46/59 active-group patients, 78.0%, and 32/66 controls, 48.5%. Infection resolution at day fifteen was recorded in 44.3% versus 31.7%, and the ordinal model favoured a better infection outcome with precision treatment, odds ratio for a worse category 0.59 (95% CI 0.38–0.91). Both findings were secondary and require interpretation alongside missing serial sampling and multiple comparisons.

There were 1,069 serious treatment-emergent events across 245 patients, 88.8% of the analysed population. This event count should not be confused with 1,069 affected individuals. Most events were not judged treatment-related. The report nevertheless identified more anaemia events with anakinra and more haemorrhage with interferon, alongside some nonserious laboratory abnormalities. Thirteen serious reactions were considered possibly or probably related across the study groups. These data describe a very ill cohort and do not establish the absence of drug-specific harm.^{1,2}

Critique

The central achievement is execution of a masked, multicountry, biomarker-guided sepsis trial with a positive primary endpoint. The investigators linked a diagnostic process to two different treatments while preserving blinding through a double-dummy design. This is more informative than an uncontrolled series showing that a biomarker normalises after treatment, and more prospective than finding a favourable subgroup after a broad trial has finished.

The primary result is also sizeable and reasonably consistent with the adjusted analysis. It should not be dismissed merely because mortality was inconclusive in a study not powered for that outcome. A phase 2 trial can appropriately use an intermediate

endpoint to decide whether a strategy deserves a larger test. The question is what that endpoint establishes, not whether every useful early trial must first prove a survival advantage.

SOFA improvement remains a surrogate. A treatment-induced change in a score is not automatically a validated substitute for survival, functional recovery or time free from support. The relationship between spontaneous SOFA improvement and better prognosis does not prove that moving the score with a drug produces the same downstream benefit. The lack of a clear ninety-day survival separation makes that distinction particularly relevant here.

The endpoint's construction adds further nuance. Averaging days two to nine captures a trajectory and reduces reliance on one day's fluctuation. Dichotomising that average at 1.4 points then discards information: patients just above and below the threshold are classified differently despite very similar courses. Baseline severity also affects the room available for improvement. The adjusted analysis is reassuring, but does not convert this particular responder definition into a universally established clinical target.

Assigning twenty-four after death avoids the serious bias of analysing only survivors' measured scores. It also gives early deaths more days at the maximum score than later deaths. The endpoint therefore weights the timing of death as well as organ dysfunction. Carrying forward discharge values makes another assumption: that the last observed state reasonably represents the unmeasured days. The discharge-exclusion sensitivity analysis supports the finding, but changes the population and does not test every possible trajectory after discharge.

These choices do not invalidate randomisation. They define what was compared. A clear account should therefore say that the strategy increased the proportion meeting the specified SOFA-based response, rather than implying a uniform physiological improvement in every treated patient or a demonstrated reduction in permanent organ injury. The heat maps are descriptive illustrations of that result, not additional independent evidence.

The combined intervention also requires a precise interpretation. Patients were randomised to receive the treatment selected by their immune classification or placebo. They were not randomised between anakinra and interferon, nor between biomarker-guided and indiscriminate use of the same drugs. ImmunoSep therefore supports the tested strategy in selected patients, but does not prove that the biomarkers are necessary to obtain benefit or that the two drugs are equally effective.

Most participants had immunoparalysis, so the overall result is weighted towards the interferon pathway. Only forty-eight had macrophage activation-like syndrome. The favourable estimates in both subgroups are encouraging, but their statistical significance does not constitute two separately powered confirmatory trials. Nor can the larger absolute difference in the small anakinra subgroup establish that this phenotype responds better. A formal comparison would need adequate information and a prespecified interaction question.

The classifications are useful operational definitions, not an exhaustive map of sepsis immunity. Ferritin is influenced by several inflammatory and tissue-injury processes. Low monocyte HLA-DR reflects an aspect of antigen presentation, but cannot capture every form of immune suppression. Patients can have both abnormalities. The rule that high ferritin takes priority is a treatment algorithm, rather than proof that hyperinflammation and immunoparalysis are biologically exclusive. The accompanying editorial appropriately stresses the limitations of dichotomous cut points.³

More than half of screened patients were unclassified. That does not establish that their immune responses were normal, nor that they could not benefit from another targeted treatment. It means the chosen tests did not place them in either eligible category. This limits the strategy's reach and prevents the result from being generalised to all patients meeting Sepsis-3 criteria.

Immune states can also change. Classification occurred near enrolment, while treatment could continue for fifteen days. Serial measurements were collected, but the trial did not establish an adaptive treatment rule that changed drug or dose whenever the phenotype changed. A future strategy might use repeated testing, yet that would be a different intervention requiring evaluation. The present findings cannot be used to infer an optimal duration or switching schedule.

The immune-reversal analysis is biologically supportive but selected. Fewer than half the cohort had the serial data used in that comparison, and sustained normalisation required repeated samples. Survival, discharge, site logistics and illness trajectory can all influence availability. Moreover, the two phenotypes used different definitions of reversal. A favourable combined percentage cannot establish that measured immune recovery mediated the SOFA effect across the whole randomised population.²

Multiplicity matters beyond the primary endpoint. Several secondary outcomes, two immune strata, exploratory time-to-event analyses and data-derived severity thresholds were examined. The post hoc cut points used the same dataset to identify higher-risk groups and assess apparent treatment interactions. Such findings are especially

vulnerable to optimism and should generate hypotheses rather than define which severity score should trigger prescribing.

Safety assessment also needs to resist broad reassurance. Nearly nine in ten patients experienced a serious treatment-emergent event, mostly reflecting their underlying illness. Relatedness judgements were masked but remain clinical attributions, and the small anakinra sample provides limited precision. The anaemia and haemorrhage signals may be chance findings among many comparisons, yet they warrant attention in a confirmatory trial. A favourable number needed to achieve the surrogate response should not be treated as directly commensurable with heterogeneous counts of serious events.

Generalisation is constrained by infection source, laboratory access and patient characteristics. Abdominal sepsis, urinary sepsis and several immunocompromised populations were excluded. Most enrolment occurred in Greece despite the six-country network. Standardised ferritin and HLA-DR testing, prompt sample transport and specialist laboratory interpretation were integral to the strategy. Demonstrating feasibility in that network does not ensure equivalent turnaround or classification in every ICU.

The appropriate next inference is therefore substantial but bounded. ImmunoSep shows that a particular precision approach can be delivered and can improve a prespecified organ-dysfunction response. Confirmation should focus on outcomes patients experience directly, with enough information to assess each treatment pathway and its harms. Routine implementation would otherwise move from a promising intermediate result to a broader clinical claim that the trial did not establish.

Body of Evidence

The 1997 confirmatory interleukin-1 receptor-antagonist trial illustrates the earlier broad-population approach. Among 696 patients with severe sepsis, twenty-eight-day mortality was 33.1% with treatment and 36.4% with placebo, without a clear difference; the study stopped after an interim assessment of futility. This did not prove that every biologically selected subgroup would fail to benefit, but it did not support unselected administration.⁴

A later reanalysis of a prior phase 3 dataset identified forty-three patients with concurrent hepatobiliary dysfunction and disseminated intravascular coagulation, interpreted as features resembling macrophage activation syndrome. Survival was higher with anakinra in that subgroup. The observation helped motivate biological

enrichment, but it was retrospective and used clinical features different from ImmunoSep's ferritin threshold. It cannot independently validate that exact cutoff.⁵

PROVIDE explored classification using ferritin and monocyte HLA-DR and linked the selected immune states to anakinra or interferon treatment. Its report described an early signal of survival with SOFA improvement, supporting further study of the approach. It provided the developmental basis for ImmunoSep, while the small treatment experience and limited duration of the early signal left the need for a larger prospective comparison.⁶

Biomarker-guided immune stimulation also has precedent beyond interferon. A thirty-eight-patient trial selected patients with persistently low monocyte HLA-DR and compared granulocyte-macrophage colony-stimulating factor with placebo. It restored HLA-DR expression in all nineteen treated patients versus three controls and suggested some clinical improvement. Its primary biological endpoint and small sample did not establish a survival benefit, but demonstrated that immune dysfunction could be used to select and monitor a randomised intervention.⁷

The 2023 interferon-gamma pneumonia-prevention trial provides an important caution. It enrolled mechanically ventilated patients with organ failure without restricting entry to ImmunoSep's immunoparalysis phenotype. The trial stopped early for potential harm after 109 randomisations; pneumonia or death occurred in 47.3% versus 30.2% of those with outcome data, adjusted hazard ratio 1.76 (95% CI 0.94–3.29). The comparison did not establish benefit and raised safety concerns.⁸

Different eligibility and timing may explain part of the contrast, but that explanation remains an inference rather than proof that biomarker selection eliminates harm. Together, these studies support prospective testing of treatment-response biology while warning against treating an immune stimulant or suppressant as universally beneficial. ImmunoSep advances that programme with a positive masked phase 2b result; it leaves patient-centred effectiveness and the full safety balance for confirmation.

Should ferritin- and HLA-DR-guided anakinra or interferon gamma become routine sepsis treatment?

ImmunoSep supports further testing of this strategy through improved early organ dysfunction. It does not yet establish a survival benefit or justify routine use beyond the studied setting.

References

1. Giamarellos-Bourboulis EJ, Kotsaki A, Kotsamidi I, Efthymiou A, Koutsoukou V, Ehler J, et al. Precision Immunotherapy to Improve Sepsis Outcomes: The ImmunoSep Randomized Clinical Trial. *JAMA*. 2026;335(9):775-786. First published online in 2025.
2. Giamarellos-Bourboulis EJ, Kotsaki A, Kotsamidi I, Efthymiou A, Koutsoukou V, Ehler J, et al. Supplementary methods and results to Precision Immunotherapy to Improve Sepsis Outcomes: The ImmunoSep Randomized Clinical Trial. *JAMA*. Published online 2025.
3. Angus DC. Precision Therapy for Sepsis: The End of the Beginning? *JAMA*. 2026;335(9):761-762. First published online in 2025.
4. Opal SM, Fisher CJ Jr, Dhainaut JF, Vincent JL, Brase R, Lowry SF, et al. Confirmatory interleukin-1 receptor antagonist trial in severe sepsis: a phase III, randomized, double-blind, placebo-controlled, multicenter trial. The Interleukin-1 Receptor Antagonist Sepsis Investigator Group. *Crit Care Med*. 1997;25(7):1115-1124.
5. Shakoory B, Carcillo JA, Chatham WW, Amdur RL, Zhao H, Dinarello CA, et al. Interleukin-1 Receptor Blockade Is Associated With Reduced Mortality in Sepsis Patients With Features of Macrophage Activation Syndrome: Reanalysis of a Prior Phase III Trial. *Crit Care Med*. 2016;44(2):275-281.
6. Leventogiannis K, Kyriazopoulou E, Antonakos N, Kotsaki A, Tsangaris I, Markopoulou D, et al. Toward personalized immunotherapy in sepsis: The PROVIDE randomized clinical trial. *Cell Rep Med*. 2022;3(11):100817.
7. Meisel C, Schefold JC, Pschowski R, Baumann T, Hetzger K, Gregor J, et al. Granulocyte-macrophage colony-stimulating factor to reverse sepsis-associated immunosuppression: a double-blind, randomized, placebo-controlled multicenter trial. *Am J Respir Crit Care Med*. 2009;180(7):640-648.
8. Roquilly A, Francois B, Huet O, Launey Y, Lasocki S, Weiss E, et al. Interferon gamma-1b for the prevention of hospital-acquired pneumonia in critically ill patients: a phase 2, placebo-controlled randomized clinical trial. *Intensive Care Med*. 2023;49(5):530-544.

SONIA

Lucinde RK, Gathuri H, Mwaniki P, Orindi B, Otieno EO, Mwakio S, et al. A Pragmatic Trial of Glucocorticoids for Community-Acquired Pneumonia. *N Engl J Med*. 2025;393(22):2187-2197.

Introduction

A treatment for pneumonia may have a different practical meaning in a hospital where intensive care is readily available and one where most critically ill patients remain on general wards. Differences in diagnostic testing, oxygen delivery, nursing capacity and access to organ support can change both baseline mortality and the consequences of treatment. Evidence generated in an ICU cannot simply be assumed to transfer unchanged to every ward population.

Corticosteroids are attractive in this setting because they are familiar, inexpensive and potentially capable of reducing inflammatory lung injury. However, their benefit in community-acquired pneumonia has varied across trials, and immunosuppression, hyperglycaemia and unrecognised alternative diagnoses complicate use. Much of the strongest mortality evidence came from well-resourced ICUs, while patients in sub-Saharan Africa were under-represented.

SONIA addressed that gap directly. It tested a short course of locally available glucocorticoids added to usual pneumonia treatment in Kenyan public hospitals, using a clinical diagnosis and a mortality endpoint. Its central contribution is therefore contextual as well as pharmacological: it asks whether an accessible treatment policy helps patients managed under the diagnostic and therapeutic conditions in which it would actually be used.¹

Synopsis

SONIA was a pragmatic, open-label, individually randomised trial in a network of eighteen Kenyan public hospitals. Patients were recruited from general medical wards, with very limited access to intensive care. Recruitment ran from April 2022 to June 2024. The study was supported by Wellcome and other public and philanthropic funders, who had no role in trial design, analysis or reporting. Written consent was obtained from patients or legally acceptable representatives.¹

Eligible adults were enrolled within forty-eight hours of hospital admission with a clinical diagnosis of community-acquired pneumonia. The definition required at least two of cough, fever, dyspnoea, haemoptysis, chest pain or crackles, present for less than fourteen days. Radiographic confirmation, pathogen testing and a formal pneumonia-severity score were not required. This reflected what the participating hospitals could routinely provide.

Patients were excluded if they had a clear indication or contraindication to corticosteroids, hospital-acquired pneumonia, pregnancy or breastfeeding. Known or suspected conditions already warranting steroids included asthma, selected obstructive-airway presentations, adrenal insufficiency and *Pneumocystis pneumonia*. Confirmed COVID-19, or a strong clinical suspicion prompting steroid treatment, also excluded participation. The trial therefore addressed uncertainty about adjunctive treatment rather than withholding steroids from patients in whom clinicians already considered them necessary.^{1,2}

An independent statistician prepared allocation cards in opaque sealed envelopes, distributed in batches and opened sequentially after eligibility and enrolment were confirmed. Patients were assigned equally to standard care plus glucocorticoids or standard care alone. Standard treatment remained the responsibility of local clinicians, usually including a beta-lactam and a macrolide. Trial-employed clinical officers recruited and followed patients but did not direct their routine treatment.

Within the active group, a second random allocation selected one of five locally available daily regimens: dexamethasone 6 mg, hydrocortisone 160 mg, methylprednisolone 30 mg, prednisolone 50 mg or prednisone 50 mg. These were the doses regarded by the trial as approximately equivalent. Treatment was intended for ten days, continuing after discharge, without tapering. Oral administration was preferred; intravenous treatment could be used initially if swallowing or enteral administration was not possible.^{1,2}

Hydrocortisone and prednisone allocation was stopped early in recruitment because the available formulations imposed a high pill burden. Most patients therefore received dexamethasone, methylprednisolone or prednisolone. The trial supplied the assigned drugs without charge. Patients were assessed daily in hospital, then contacted by telephone on days fourteen and thirty. Postdischarge adherence was based on patient or family report.

The primary outcome was all-cause death within thirty days of enrolment, including deaths after discharge. Secondary outcomes assessed mortality at seven, fourteen and twenty-one days and separated in-hospital from out-of-hospital deaths. Adverse events

and serious adverse events were recorded. A prespecified immune-response outcome was collected but was not included in this publication. An independent monitoring board reviewed an interim analysis and recommended that recruitment continue.

The target of 2,180 patients provided 85% power to detect a 25% relative mortality reduction, from 20% to 15%, allowing for 5% loss to follow-up. The primary intention-to-treat analysis used a Cox model stratified by hospital. Participants with incomplete follow-up were censored at their last observation. Prespecified subgroup analyses examined demographic, regional, oxygenation and treatment characteristics; secondary comparisons were not adjusted for multiplicity. Two post hoc sensitivity analyses excluded incomplete cases or patients who had received steroids before randomisation.^{1,2}

Of 46,224 medical-ward admissions screened, 2,180 patients were randomised: 1,089 to glucocorticoids and 1,091 to standard care. Most exclusions reflected failure to meet the clinical pneumonia definition. Median age was 53 years and 46.3% were women. Oxygen saturation was below 90% in 37.1%, and systolic pressure below 90 mmHg in approximately 8%. Only five participants were transferred to intensive care during the admission.¹

HIV infection was known in 344 patients, 15.8%, but status was unknown in approximately 46%. Hypertension and diabetes were other common comorbidities. Chest radiographs were available for only 39%, without formal radiologist reports. Thirty-one patients had known pulmonary tuberculosis at baseline. The cohort therefore included diagnostic uncertainty and comorbidity patterns that differed materially from many previous pneumonia trials.

Within the active group, 371 patients received dexamethasone, 352 methylprednisolone and 343 prednisolone; twelve received hydrocortisone and eleven prednisone. Fifteen initially received an intravenous formulation. Trial steroids began within a day of enrolment, although exact timing in hours was not captured. Median inpatient steroid exposure was four days and median hospital stay five days. Reported medication adherence was high, but the remaining postdischarge course was not directly observed.²

Thirty-day vital status was known for 2,082 patients, 95.5%. Forty-three active-group and fifty-five control patients had incomplete outcomes: seventy-five had withdrawn and twenty-three were lost to follow-up. Reasons included a new diagnosis or a new indication for corticosteroids. Thirty-three controls and sixteen active-group patients withdrew after receiving glucocorticoids for another medical condition.

There were 246 reported deaths among 1,089 patients assigned glucocorticoids, 22.6%, and 284 among 1,091 controls, 26.0%. The site-stratified hazard ratio was 0.84 (95% CI 0.73–0.97; $P=0.02$). The raw difference was approximately 3.4 percentage points, while the survival model incorporated follow-up time and censoring. In-hospital deaths numbered 196 versus 223, and deaths outside hospital fifty versus sixty-one.¹

Earlier mortality estimates pointed in the same direction. The complete-case sensitivity analysis gave a hazard ratio of 0.84 (95% CI 0.73–0.96); excluding thirty-eight patients who had received steroids before randomisation gave 0.83 (95% CI 0.72–0.97). The trial was not powered to determine whether effects differed reliably by individual steroid, hypoxaemia or other subgroup characteristics.²

Any adverse event affected 184 active-group patients, 16.9%, and 154 controls, 14.1%. Serious adverse events affected forty-four, 4.0%, and forty-eight, 4.4%, respectively. Five patients, 0.5% of the steroid group, had serious events judged possibly related to treatment. Hyperglycaemia accounted for thirty-five recorded adverse events in the steroid group, including four serious events. These are event counts, not evidence that 16.6% of all treated patients developed hyperglycaemia. Pulmonary tuberculosis was frequently diagnosed during follow-up in both groups.^{1,2}

Critique

The trial's most important strength is its directness for a previously neglected clinical setting. Randomisation was embedded in public hospitals caring for patients with high pneumonia mortality and little access to intensive care. It evaluated drugs and routes that were locally available, and chose death rather than an outcome dependent on advanced diagnostics or discretionary ICU admission. Follow-up after discharge captured a substantial fraction of deaths that an in-hospital-only endpoint would miss.

The primary comparison supports a mortality benefit for the tested strategy. Its size, multicentre recruitment and more than five hundred deaths provide substantial information. The confidence interval nonetheless extends close to no effect, and the observed reduction was smaller than the five-percentage-point difference anticipated in planning. The result is persuasive evidence for a context-specific policy, while remaining less precise than a headline relative reduction might imply.

The hazard ratio should not be translated into an identical percentage reduction in thirty-day risk. It compares instantaneous hazards over follow-up under the model. The reported crude proportions suggest an absolute difference of about 3.4 points, but include participants whose final vital status was unknown. An exact number needed to

treat calculated from those proportions would imply more certainty than the incomplete follow-up permits. The survival analysis is the principal estimate.

Open-label treatment is a limitation even with mortality as the endpoint. Knowledge of allocation can change surveillance, discharge decisions, treatment escalation or attention to complications. The study documented broad antibiotic patterns but did not comprehensively measure all cointerventions. Mortality is less subjective than a clinical-stability assessment, yet it can still be influenced through differences in care. Randomisation and separation of trial follow-up from routine treatment reduce some concerns without eliminating them.

The envelope system provided a practical method of allocation concealment. Its protection depended on secure storage, sequential opening and enrolling patients before revealing the card. This is distinct from the later absence of masking. A trial can have adequate concealment at allocation while remaining vulnerable to treatment-related performance differences afterwards. The report describes the intended safeguards, but cannot provide the same control over subsequent behaviour as a placebo design.

Broad clinical eligibility is both an advantage and a source of uncertainty. Requiring chest imaging or microbiological confirmation would have excluded many patients who actually receive pneumonia treatment in these hospitals. The trial therefore estimates the effect of adding steroids to a clinical pneumonia policy. It does not establish the effect in a microbiologically confirmed bacterial-pneumonia population, nor determine which pathogens or alternative diagnoses account for benefit.

Unrecognised steroid-responsive conditions may have contributed. HIV was common, status was often unknown, and *Pneumocystis pneumonia* or viral infection could be difficult to identify. Some patients developed a clearer alternative diagnosis after enrolment and withdrew. That possibility does not erase the pragmatic result: diagnostic uncertainty is part of the setting being studied. It does limit the mechanistic claim that the observed benefit must represent suppression of inflammation from typical bacterial pneumonia.

Likewise, residence on a general ward should not be taken as evidence of mild disease. More than a third of patients were hypoxaemic, approximately one in twelve hypotensive, and control mortality was 26%. Many might have received higher-level monitoring or organ support in another health system. SONIA should therefore not be used as a simple justification for steroids in all low-risk ward pneumonia in a well-resourced hospital.

The absence of formal severity scores limits attempts to locate the effect within conventional severe-pneumonia categories. Oxygen saturation supplies useful information but is not a complete severity assessment. Point estimates in individual oxygenation, sex or regional subgroups should not be interpreted as definitive effect modification merely because one interval excludes one and another does not. The study's principal question concerns the overall enrolled population.

Missing outcomes deserve more weight than their percentage alone might suggest. Ninety-eight patients lacked final vital status, compared with thirty-eight fewer observed deaths in the steroid arm. The groups had reasonably similar attrition, and complete-case analysis was consistent, but missingness was not entirely random in its clinical causes. New diagnoses, treatment failure and new indications for steroids can be associated with prognosis. Censoring such patients requires assumptions about their unobserved subsequent outcomes.²

The supplementary analysis appropriately acknowledges that observed data cannot distinguish missing at random from missing not at random with certainty. A complete-case analysis does not solve informative loss if the excluded patients differ in unmeasured prognosis. Conversely, the presence of some missingness does not establish that the treatment effect is artefactual. The result should carry this limitation rather than being described either as fully complete or as uninterpretable.

Steroid selection adds another layer. The trial tested a policy allowing several regimens, with further randomisation within the active arm. That design does not supply equally precise evidence for every drug. Hydrocortisone and prednisone were used in very few patients before their withdrawal, and the later mixture was dominated by three agents. A pooled benefit cannot establish dose equivalence, route equivalence or superiority of one mineralocorticoid profile.

The pill-burden problem is clinically informative. A pharmacologically plausible regimen may be difficult to deliver when local tablet strengths require many pills. Stopping two formulations was an implementation adjustment rather than evidence that their biological effect was inferior. Any recommendation based on SONIA needs to preserve the distinction between a feasible local treatment policy and a claim that one specific ten-day prescription has been independently proven optimal.

Adherence also has limits. Inpatient administration was directly monitored, but postdischarge reporting asked generally whether prescribed medicines had been taken. Very high reported adherence therefore does not prove that every patient completed ten steroid doses. Median inpatient exposure of four days should not be mistaken for

the intended total treatment duration, and the study cannot determine whether four, seven or ten days would have been sufficient.

Safety interpretation requires correct denominators. The thirty-five hyperglycaemia events represented 16.6% of all adverse events recorded in the steroid arm, not 16.6% of steroid recipients. Similarly, four serious hyperglycaemia events were 8.9% of that arm's serious-event count. Neither percentage is a patient-level incidence. The overall serious-event proportions were close, but ascertainment in a resource-limited setting may miss metabolic or infectious complications that would be detected by more intensive monitoring.²

The editorial's emphasis on glucose monitoring is therefore reasonable, even though its stated hyperglycaemia percentage should not be carried forward as a patient risk. Blood-glucose testing, recognition of alternative infection and the ability to respond to deterioration are part of safe implementation. They are practical requirements arising from the observed intervention, not reasons to assume that a low-cost drug has no meaningful service implications.³

Longer-term effects remain uncertain. Follow-up ended at thirty days, which captures the main acute mortality question but not every consequence of immunosuppression or tuberculosis progression. Similar numbers of tuberculosis diagnoses over this short interval do not establish long-term infectious safety. Nor did the trial provide a formal cost-effectiveness analysis: free study medication and dedicated follow-up staff may differ from routine delivery.

SONIA's result should thus be used in the population and circumstances it actually studied. It provides important evidence that a feasible adjunctive steroid strategy can reduce short-term mortality among clinically diagnosed pneumonia patients in Kenyan wards with high baseline risk. It does not remove the need to identify existing indications, contraindications or alternative diagnoses, and it does not make ICU trials from different settings redundant.

Body of Evidence

The Swiss STEP trial randomised 785 hospitalised patients to prednisone 50 mg daily for seven days or placebo. Median time to clinical stability was three versus 4.4 days, while hyperglycaemia requiring insulin occurred more often with prednisone. Its principal contribution concerned speed of recovery rather than a mortality-powered comparison. SONIA extended the ward-based question to a population with much higher risk of death and different access to rescue treatment.⁴

ESCAPE tested a prolonged methylprednisolone course in severe pneumonia across United States Veterans Affairs centres. It enrolled 586 patients, well below its intended target, and analysed 584. Sixty-day mortality was 16% versus 18%, adjusted odds ratio 0.90 (95% CI 0.57–1.40). The uncertain result does not prove methylprednisolone ineffective, but shows why drug, timing, duration and population cannot be collapsed into a single uniform steroid effect.⁵

CAPE COD provided stronger mortality evidence in French ICUs. Among 795 analysed patients, intravenous hydrocortisone reduced twenty-eight-day mortality from 11.9% to 6.2%. Treatment used 200 mg daily initially, followed by a response-dependent course and taper. Its patients had access to intensive monitoring and organ support, and its regimen differed from SONIA's predominantly oral, fixed-duration policy. The studies support benefit in different settings without being direct replications.⁶

The 2025 REMAP-CAP corticosteroid report did not establish a survival benefit from fixed hydrocortisone 50 mg every six hours for seven days in severe non-COVID pneumonia. That arm stopped for futility regarding a large mortality reduction, while smaller benefits and harm remained possible. It should not be cited as another positive hydrocortisone mortality trial. Its adaptive platform design and more heterogeneous ICU population are discussed separately in this book.⁷

RECOVERY supplied a rationale for SONIA's dose and duration, using dexamethasone 6 mg daily for up to ten days in COVID-19. Benefit depended on respiratory-support status, with no established benefit in patients receiving no support. That pathogen-specific result was useful for trial design but did not itself justify treating clinically diagnosed non-COVID pneumonia. SONIA supplied the direct randomised evidence for its own setting.⁸

The overall literature therefore supports a differentiated interpretation. Steroids can improve important outcomes in some pneumonia populations, but positive and inconclusive trials coexist. SONIA adds a large, locally conducted mortality trial in a setting poorly represented by earlier studies. Its value lies in answering that real clinical question, while preserving uncertainty about the best agent, the patients most likely to benefit and the effect in lower-risk or differently resourced populations.

Should SONIA change adjunctive steroid treatment for pneumonia in settings resembling the Kenyan trial hospitals?

It supports considering the tested short-course strategy in comparable high-risk ward populations, with glucose monitoring and attention to alternative diagnoses. It does not establish benefit for every pneumonia patient or identify a single superior steroid regimen.

References

1. Lucinde RK, Gathuri H, Mwaniki P, Orindi B, Otieno EO, Mwakio S, et al. A Pragmatic Trial of Glucocorticoids for Community-Acquired Pneumonia. *N Engl J Med*. 2025;393(22):2187-2197.
2. Lucinde RK, Gathuri H, Mwaniki P, Orindi B, Otieno EO, Mwakio S, et al. Supplementary methods and results to A Pragmatic Trial of Glucocorticoids for Community-Acquired Pneumonia. *N Engl J Med*. 2025;393(22):2187-2197.
3. Kwizera A, Dünser MW. Glucocorticoids for Pneumonia in Africa - Old Therapy, New Context. *N Engl J Med*. 2025;393(22):2263-2264.
4. Blum CA, Nigro N, Briel M, Schuetz P, Ullmer E, Suter-Widmer I, et al. Adjunct prednisone therapy for patients with community-acquired pneumonia: a multicentre, double-blind, randomised, placebo-controlled trial. *Lancet*. 2015;385(9977):1511-1518.
5. Meduri GU, Shih MC, Bridges L, Martin TJ, El-Solh A, Seam N, et al. Low-dose methylprednisolone treatment in critically ill patients with severe community-acquired pneumonia. *Intensive Care Med*. 2022;48(8):1009-1023.
6. Dequin PF, Meziani F, Quenot JP, Kamel T, Ricard JD, Badie J, et al. Hydrocortisone in Severe Community-Acquired Pneumonia. *N Engl J Med*. 2023;388(21):1931-1941.
7. The REMAP-CAP Investigators. Effect of hydrocortisone on mortality in patients with severe community-acquired pneumonia: The REMAP-CAP Corticosteroid Domain Randomized Clinical Trial. *Intensive Care Med*. 2025;51(4):665–680.
8. Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L, et al. Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med*. 2021;384(8):693-704.

REMAP-CAP: Corticosteroids

The REMAP-CAP Investigators. Effect of hydrocortisone on mortality in patients with severe community-acquired pneumonia: The REMAP-CAP Corticosteroid Domain Randomized Clinical Trial. *Intensive Care Med.* 2025;51(4):665–680.

Introduction

Corticosteroids have become an important component of treatment for selected patients with severe pneumonia, but the evidence does not describe a single uniform intervention. Trials have used different drugs, routes, durations and stopping rules, in populations ranging from general-ward patients to those requiring ventilation and vasopressors. A positive result in one setting does not settle every remaining question about severe community-acquired pneumonia.

Hydrocortisone is particularly attractive when pneumonia causes circulatory failure. Its anti-inflammatory and haemodynamic effects might improve organ function, yet faster resolution of shock does not necessarily translate into improved survival. The balance may also differ in influenza, where immunosuppression could have different consequences from those in bacterial infection. These uncertainties motivated separate assessment of patients with and without shock and influenza.

REMAP-CAP approached the question through an international adaptive platform. The corticosteroid domain evaluated several strategies while the wider platform studied other treatments for severe pneumonia. This report concerns one specific comparison: a fixed seven-day course of intravenous hydrocortisone against no routine corticosteroid. Its result challenges an assumption of a large, broadly applicable mortality benefit, while its design requires care in explaining what futility does and does not mean.¹

Synopsis

This open-label randomised comparison recruited adults at 101 sites in eighteen countries between March 2018 and December 2023. The platform had multiple international sponsors and funders, who had no role in design, analysis or reporting. An independent data and safety monitoring board reviewed accumulating results, while the international steering committee remained blinded to aggregate outcomes. Consent procedures followed local requirements, including deferred consent where permitted.¹

Eligible patients had a clinical lower respiratory infection with a new radiographic infiltrate considered infective, and required intensive care within forty-eight hours of hospital presentation. Respiratory organ support included invasive or non-invasive ventilation, or high-flow nasal oxygen at a flow of at least 30 L/min and an inspired oxygen fraction of at least 0.4. Cardiovascular support meant an intravenous vasopressor or inotrope. Randomisation in the corticosteroid domain had to occur within twenty-four hours of ICU admission.

Exclusions included recent inpatient healthcare exposure, residence in a nursing home or long-term care facility, imminent death without commitment to full support, chronic systemic corticosteroid treatment, corticosteroid hypersensitivity and known or presumed COVID-19. Later platform expansions to children and less severe pneumonia did not contribute patients to this comparison. The population was adults with severe, predominantly non-COVID community-acquired pneumonia.^{1,2}

The active strategy prescribed hydrocortisone 50 mg intravenously every six hours for seven days, without tapering. Treatment stopped earlier at hospital discharge. Controls received no routine corticosteroid, but both groups could receive systemic steroids for a new established indication, such as bronchospasm, anaphylaxis or post-extubation stridor. Other treatment followed local practice, and patients could be randomised in other platform domains.

Sites needed to offer at least two corticosteroid-domain options, but did not have to offer every arm. A site unwilling to withhold steroids could compare different steroid regimens without participating in the control group. The main analysis consequently included some patients from sites that could not randomise between these two particular options. A separate analysis restricted the comparison to sites offering both fixed-duration hydrocortisone and control.¹

Allocation began balanced between locally available arms and subsequently used response-adaptive probabilities based on accumulating outcomes. At the first adaptive analysis in August 2021, a data-transfer error reversed the treatment labels for twenty-eight patients in the analysis file. This incorrectly favoured hydrocortisone in later allocation probabilities. The actual treatments assigned to patients and the underlying clinical database were correct. The final outcome analysis used corrected data.^{1,2}

The error was discovered in February 2024, after the fixed-duration arm had closed. Closure followed a November 2023 adaptive analysis using correct data that met the prespecified futility rule; recruitment stopped on 6 December. Other corticosteroid

strategies had not met stopping criteria and were not reported in this publication. The error therefore affected subsequent allocation balance, but did not mean that the final mortality analysis knowingly used reversed treatment assignments.

The primary outcome was all-cause mortality at ninety days. A Bayesian hierarchical logistic model estimated treatment effects in four groups defined by shock and suspected or confirmed influenza. It adjusted for age, sex, site, country, calendar period, severity, eligibility and assignments in other platform domains. Data from other eligible platform participants helped estimate these covariate relationships; they did not constitute additional patients randomised to hydrocortisone versus control.

The model allowed partial borrowing of information between the four shock–influenza groups. Treatment-effect priors were centred on no effect, with interaction priors pulling subgroup differences towards a common effect. Sensitivity analyses altered this borrowing and restricted the population. The COVID-19 corticosteroid effect was modelled separately, rather than assuming that its treatment benefit applied directly to non-COVID pneumonia.^{1,2}

Superiority required a posterior probability above 99% that the mortality odds ratio was below one. The numerical futility criterion concerned an odds ratio below 1/1.2, approximately 0.83: the posterior probability of achieving an effect at least this large had to be below 5%. This is a threshold for a specified large benefit, not a test demonstrating that every possible benefit is absent.

Among 658 patients in the two reported groups, 536 were assigned hydrocortisone and 122 control. Of the hydrocortisone group, 197 came from sites not offering control; six controls came from sites not offering fixed-duration hydrocortisone. The subset with both options available comprised 455 patients at seventy sites, with 339 assigned hydrocortisone and 116 control.¹

Median age was 62.5 versus 58.5 years. Diabetes was present in 30.3% versus 22.1%, and invasive ventilation in 39.7% versus 37.7%. Approximately half had shock. Only fifty-four had suspected or confirmed influenza for the stratified analysis, including nine controls. Several baseline characteristics therefore suggested greater illness or comorbidity in the larger active-treatment group.

Treatment data were available for 533 hydrocortisone patients and all controls. Hydrocortisone was received by 527 of 533, 98.9%, with a median course of seven days. Twenty-eight controls, 23.0%, received some systemic corticosteroid during admission. Exposure among those controls was generally shorter; only one received more than six

days at a hydrocortisone-equivalent dose of at least 200 mg daily. This was crossover or rescue exposure, not wholesale delivery of the assigned active regimen to controls.²

Ninety-day vital status was available for 521 hydrocortisone patients and all 122 controls. There were seventy-eight deaths, 15.0%, versus twelve, 9.8%. All fifteen missing primary outcomes were in the hydrocortisone group. The adjusted median mortality odds ratios ranged from 1.52 to 1.63 across the four strata, with every 95% credible interval including one. Posterior probabilities of any mortality benefit ranged from 9.2% to 15.7%.¹

In a sensitivity model pooling the four strata, the adjusted odds ratio was 1.56 (95% credible interval 0.80–3.31), with a 10.0% probability of benefit and a 3.4% probability of meeting the specified large-benefit threshold. Restricting analysis to sites offering both options gave a similar pooled estimate of 1.55 (95% credible interval 0.75–3.63). These estimates favoured control but remained imprecise.

At the final analysis, the probabilities of the specified large benefit were 7.1% and 5.3% in the two influenza strata, and 4.4% and 3.3% in the corresponding non-influenza strata. The final influenza estimates therefore did not all remain below the interim 5% threshold. Analyses without borrowing were particularly uncertain for influenza, reflecting the small numbers directly available.^{1,2}

ICU mortality was 9.3% versus 5.7%, with wide uncertainty. Median ICU stay was five versus 5.5 days, and hospital stay twelve versus fourteen days; adjusted analyses did not establish improved recovery. Organ-support-free days showed some probability of benefit, but did not meet the trial's superiority criterion. An exploratory analysis found a median shock duration of two versus three days, with $P=0.05$. Serious adverse events were reported in seven hydrocortisone patients and one control.¹

Critique

The trial addressed a clinically important question across a broad international network and used a mortality endpoint extending beyond hospital discharge. Its inclusion of patients with and without shock is valuable, particularly because some influential pneumonia trials excluded septic shock. Recruitment in diverse health systems also tests whether benefit persists beyond a small group of highly selected centres.

The most defensible conclusion is that this fixed-duration regimen was unlikely to produce the large survival advantage specified by the platform. Futility should not be translated into proof of equivalence. The pooled credible interval still included a

clinically relevant reduction in mortality odds, as well as substantial harm. A stopping rule can reasonably conclude that further recruitment is unlikely to establish the target benefit while leaving uncertainty about smaller effects.

Conversely, the direction of the mortality estimates cannot be ignored. Both crude mortality and adjusted estimates favoured control, and posterior probabilities of harm were high. Those findings argue against describing the study as simply neutral. They do not establish that hydrocortisone caused excess mortality: the interval was wide, the control group small, and the reported inferiority rule was not the same as a conventional threshold of 95% probability of harm.

The mathematical scale matters. The published shorthand about a large relative improvement can be mistaken for a 20% reduction in absolute risk, relative risk or mortality odds. The actual numerical threshold was an odds ratio below approximately 0.83. An odds ratio and a risk ratio are not interchangeable, and the meaning of an odds reduction depends on baseline risk. Stating the threshold explicitly avoids overstating what the futility finding excludes.

The allocation error is a material limitation, although its consequences need accurate description. It did not involve clinicians administering the wrong assigned treatment, nor were the final outcome labels left incorrect. It influenced the adaptive algorithm, increasing later assignment to hydrocortisone and reducing the information collected in controls. Correcting the final dataset could repair the analysis labels but could not retrospectively restore the more balanced sample that a different allocation history might have produced.

Response-adaptive allocation can itself generate unequal groups even without error. Here, however, the erroneous early signal is particularly relevant because only twelve observed deaths occurred in the control arm. A large total platform enrolment should not distract from that limited direct comparison. Modelled adjustment and borrowing improve use of available information, but cannot create the same evidence as a much larger independently randomised control group.

Site participation introduces another distinction from a conventional two-arm trial. Nearly two hundred hydrocortisone patients were enrolled where control was unavailable. Their inclusion requires modelling to account for site, time and other differences. The analysis restricted to sites offering both options is therefore especially important. Its similar point estimate reduces concern that the whole result was driven by non-comparable sites, but its wide interval preserves substantial uncertainty.

Baseline differences reinforce the need for adjustment. Hydrocortisone patients were older and more often diabetic, with indications of greater circulatory severity. The primary model included important prognostic variables, but adjustment cannot guarantee removal of every imbalance, particularly in a modest comparison spread across many sites. Correspondence after publication highlighted baseline severity and control-group steroid exposure as interpretative concerns. These are reasons for caution, rather than evidence that an unobserved large benefit must have been present.³

Control exposure complicates the estimand. The trial compared a policy of routine seven-day hydrocortisone with no routine steroid while allowing treatment for new indications. It did not compare seven days of steroid with a guarantee of no steroid exposure at any time. Rescue treatment could reduce separation, but the brief courses received by some controls were not equivalent to universal crossover. The intention-to-treat result remains relevant to a policy that preserves clinically necessary rescue treatment.

The open-label design permits differences in cointerventions, investigation and discharge decisions. Death is less open to subjective classification than a measure of clinical improvement, but treatment decisions can still affect it. Secondary outcomes such as shock duration and length of stay are more susceptible to practice variation. They should not be used to overturn an uncertain mortality result, especially when the exploratory shock comparison was borderline and not protected against multiple testing.

The hierarchical model also deserves clinical interpretation. Partial pooling assumes that extreme differences between the four strata are less likely than broadly related effects, while allowing the data to modify that assumption. This is useful when subgroup numbers are small, but means that an influenza estimate reflects information beyond the few influenza patients themselves. It is not an independent adequately powered influenza trial.

Without borrowing, the influenza intervals became extremely wide and allowed both marked benefit and harm. The final influenza large-benefit probabilities also rose above the interim stopping threshold. Neither observation invalidates the platform's stopping decision, which used the information available at that time. Both prevent a confident claim that this study settled steroid treatment for severe influenza pneumonia. Sparse subgroup evidence should remain visibly sparse.

Missing primary outcomes were few overall but confined to hydrocortisone. Fifteen unobserved outcomes are not negligible beside twelve control deaths. The observed

mortality difference cannot be treated as if follow-up were complete, and model-based estimates depend on how missingness relates to prognosis. This limitation adds uncertainty without, by itself, specifying whether the missing patients would have strengthened or weakened the apparent disadvantage.

Safety reporting was similarly limited. Seven versus one serious adverse events cannot establish equal safety, and a low recorded count does not imply absence of every metabolic, infectious or neuromuscular complication. Some clinically important harms may be reflected in outcomes rather than classified as treatment-related events. The trial is more informative about mortality under the treatment policy than about precise rates of individual steroid toxicities.

Finally, the result belongs to the tested regimen. Hydrocortisone boluses for seven days without tapering are not identical to a response-dependent infusion, a longer methylprednisolone course, or hydrocortisone combined with fludrocortisone. Differences between studies may reflect population, delivery, timing, chance or combinations of these factors. Cross-trial comparison can generate explanations but cannot identify which difference caused the contrasting findings.

Body of Evidence

CAPE COD randomised patients with severe community-acquired pneumonia in French ICUs to hydrocortisone or placebo. Among 795 analysed patients, twenty-eight-day mortality was 6.2% versus 11.9%. Hydrocortisone began at 200 mg daily, with subsequent duration and taper guided by clinical response. Septic shock at enrolment was excluded. Its positive result is directly relevant, but it evaluated a different patient mix and treatment schedule from REMAP-CAP.⁴

ESCAPE studied a longer methylprednisolone regimen in severe pneumonia, enrolling 586 patients against a substantially larger planned target. In the 584 analysed patients, sixty-day mortality was 16% versus 18%, adjusted odds ratio 0.90 (95% CI 0.57–1.40). This uncertain result already showed that a plausible steroid strategy need not demonstrate a mortality benefit, and that imprecision should not be mistaken for equivalence.⁵

Septic-shock trials provide another relevant comparison. ADRENAL randomised 3,800 ventilated patients to hydrocortisone 200 mg daily or placebo. Ninety-day mortality was 27.9% versus 28.8%, without a significant difference, although shock resolved faster. It supports separating haemodynamic improvement from a proven survival effect, a distinction also relevant to REMAP-CAP's exploratory shock-duration finding.⁶

An a priori planned exploratory APROCCHSS subgroup analysis examined 562 patients with community-acquired-pneumonia-related septic shock. Ninety-day mortality was 39% with hydrocortisone plus fludrocortisone and 51% with placebo, odds ratio 0.60 (95% CI 0.43–0.83). This is evidence from a randomised parent trial, but remains a subgroup analysis of a combined regimen. It neither isolates the effect of hydrocortisone nor proves that adding fludrocortisone explains the difference from REMAP-CAP.⁷

SONIA subsequently randomised 2,180 adults with clinically diagnosed pneumonia in Kenyan general wards, where access to intensive care was very limited. A predominantly oral, ten-day glucocorticoid strategy reduced thirty-day mortality, hazard ratio 0.84 (95% CI 0.73–0.97). High baseline mortality, limited diagnostic testing and a mixture of steroid agents make it an important complementary trial rather than a direct replication of an ICU hydrocortisone regimen.⁸

Together, these studies support a differentiated account of corticosteroids in pneumonia. Some strategies have improved survival in their enrolled populations, while others have not established benefit. REMAP-CAP adds a caution against assuming a large mortality reduction from routine fixed-duration hydrocortisone across all severe non-COVID pneumonia. It leaves room for smaller effects, population differences and other regimens, while making those uncertainties explicit.

Does REMAP-CAP support routine fixed seven-day hydrocortisone for all adults with severe community-acquired pneumonia?

It makes a large mortality benefit from this regimen unlikely, while leaving smaller benefit and harm unresolved. Other steroid regimens and separate indications require their own evidence.

References

1. The REMAP-CAP Investigators. Effect of hydrocortisone on mortality in patients with severe community-acquired pneumonia: The REMAP-CAP Corticosteroid Domain Randomized Clinical Trial. *Intensive Care Med.* 2025;51(4):665–680.
2. The REMAP-CAP Investigators. Supplementary methods and results to Effect of hydrocortisone on mortality in patients with severe community-acquired pneumonia. *Intensive Care Med.* 2025;51:665–680.

3. Dequin PF, Giraudeau B, Le Gouge A. Correspondence about the article "Effect of hydrocortisone on mortality in patients with severe community-acquired pneumonia. The REMAP-CAP corticosteroid domain randomized clinical trial". *Intensive Care Med.* 2025;51(9):1740-1741.
4. Dequin PF, Meziani F, Quenot JP, Kamel T, Ricard JD, Badie J, et al. Hydrocortisone in Severe Community-Acquired Pneumonia. *N Engl J Med.* 2023;388(21):1931-1941.
5. Meduri GU, Shih MC, Bridges L, Martin TJ, El-Solh A, Seam N, et al. Low-dose methylprednisolone treatment in critically ill patients with severe community-acquired pneumonia. *Intensive Care Med.* 2022;48(8):1009-1023.
6. Venkatesh B, Finfer S, Cohen J, Rajbhandari D, Arabi Y, Bellomo R, et al. Adjunctive Glucocorticoid Therapy in Patients with Septic Shock. *N Engl J Med.* 2018;378(9):797-808.
7. Heming N, Renault A, Kuperminc E, Brun-Buisson C, Megarbane B, Quenot JP, et al. Hydrocortisone plus fludrocortisone for community acquired pneumonia-related septic shock: a subgroup analysis of the APROCCHSS phase 3 randomised trial. *Lancet Respir Med.* 2024;12(5):366-374.
8. Lucinde RK, Gathuri H, Mwaniki P, Orindi B, Otieno EO, Mwakio S, et al. A Pragmatic Trial of Glucocorticoids for Community-Acquired Pneumonia. *N Engl J Med.* 2025;393(22):2187-2197.

ENCOMPASS

Taylor SP, Eaton T, Rios A, Boyd D, Tapp H, McWilliams A, et al. Proactive Telehealth-Based Sepsis Transition and Recovery Support, Hospital Readmission, and Mortality: A Randomized Clinical Trial. JAMA Intern Med. 2025;185(10):1238-1246.

Introduction

Surviving the acute phase of sepsis does not end the illness. Patients may leave hospital with new disability, altered medication regimens and unstable chronic disease, while responsibility for their care transfers between several services. Telephone support offers a practical way to identify problems early and coordinate treatment without requiring every patient to attend a specialist clinic. Its usefulness, however, depends on whether contact leads to timely, effective care and whether the patients most in need can be reached.

The sepsis transition and recovery programme, STAR, uses nurse navigators to support this vulnerable period. Its components include medication review, assessment of new physical or cognitive problems, monitoring of conditions likely to deteriorate, and palliative care when appropriate. An earlier individually randomised study, IMPACTS, suggested that thirty days of STAR support could reduce the combined risk of death or readmission. ENCOMPASS tested whether this approach remained effective when extended to ninety days and implemented across a broader range of hospitals.^{1,2}

The choice of endpoint creates an additional question. Readmission is burdensome and sometimes preventable, but returning promptly to hospital may also be the correct response to a dangerous complication. A programme that identifies deterioration could increase appropriate readmissions while preventing deaths. Combining the two outcomes assumes more alignment than may exist in practice. That possibility does not justify changing the primary endpoint after seeing the results, but it matters when deciding what a neutral composite actually means.³

ENCOMPASS therefore addresses both the effectiveness of a scalable recovery service and the difficulty of measuring recovery. Its results are particularly relevant to systems considering centralised navigation across geographically dispersed hospitals.

Synopsis

ENCOMPASS was a pragmatic stepped-wedge cluster randomised trial within the Atrium Health system in North Carolina. Recruitment ran from July 2020 to June 2023, with follow-up through December 2023. Eight hospitals were initially assigned a transition sequence, but one withdrew after leaving the health system and before implementing STAR. Seven hospitals contributed to the analysis, including urban, suburban and rural community hospitals. Investigators disclosed National Institutes of Health support, including support from the National Institute of Nursing Research. An institutional review board waived individual research consent, and a monitoring board oversaw the trial. Patients could decline the offered clinical service.²

Every hospital began with usual care. At four-month intervals, hospitals moved to STAR in a randomly determined order. The sequence was concealed from each hospital until one month before its scheduled transition, allowing staff training and preparation. Nine four-month periods covered thirty-six months. Clinicians and investigators were unmasked, but the principal outcomes came from routinely collected records.

An electronic algorithm identified adults admitted through the participating hospital's emergency department. Eligibility required suspected infection, defined by antibiotic and bacterial culture orders with at least two systemic inflammatory response markers; organ dysfunction, represented by at least two points on the admission SOFA or quick-SOFA assessment; and a predicted ninety-day readmission risk of at least 25%. Patients had to remain in hospital when the weekday eligibility list was generated. Exclusions included an early change in resuscitation status, residence more than two-and-a-half hours away, participation in another care-management programme and previous trial enrolment.^{2,4}

Five specialist nurse navigators delivered STAR centrally. They reviewed records and contacted patients, carers and inpatient teams before discharge. The service included attention to ongoing infection, discharge planning, medication reconciliation, new impairments, symptoms and access to follow-up. After discharge, navigators aimed to contact patients within forty-eight hours and continued telephone support for up to ninety days. The main report describes calls every three to five days during the first month, followed by longer intervals; the supplementary workflow describes further tailoring according to ongoing risk.

The service was more than a telephone reminder. Positive symptom screens could prompt contact with primary care or the hospital transition service, with escalation to a clinician assessment. Navigators could arrange medication review, identify barriers to appointments, coordinate community support and encourage goals-of-care discussions.

Patients discharged to facilities as well as those returning home could receive support. Patients and navigators could agree that further contact was unnecessary. Control periods retained usual hospital and postdischarge care without access to STAR.⁴

The primary endpoint was all-cause death or hospital readmission assessed through ninety days after discharge. Its detailed definition included deaths during the index admission, because enrolment occurred before discharge. Readmission included inpatient and observation encounters at any Atrium Health hospital. Death ascertainment combined electronic records with national death data. Secondary outcomes included the separate composite components, cause-specific readmissions, days alive outside hospital, emergency department attendance and outpatient visits. Patient-reported quality of life was not collected.^{2,4}

The original target of 4,032 patients assumed eight hospitals and at least 90% power to detect an eight-percentage-point absolute reduction in the primary endpoint. After one hospital withdrew, a revised calculation indicated that 3,416 participants across seven hospitals would retain at least 80% power for that difference. The final sample exceeded this revised target.

The primary intention-to-treat analysis used generalised estimating equations with a binomial distribution and logit link, accounting for hospital clustering. Adjustment included age, baseline readmission risk, initial SOFA score, COVID-19 diagnosis and implementation period. Investigators changed from the initially preferred mixed-effects approach after finding little hospital-level random-effect variation and difficulties fitting the correlation structure. The monitoring board approved the approach; the mixed model was retained as a sensitivity analysis. Secondary comparisons were not adjusted for multiplicity. A prespecified modified analysis excluded patients who died during the index admission or left against medical advice.^{2,4}

Among 19,151 patients screened, 3,548 were enrolled: 2,122 during STAR periods and 1,426 during usual-care periods. Median age was 68 years, 52% were women and almost all had at least one recorded comorbidity. One-third required intensive care. Median presenting SOFA score was 4, median predicted readmission risk was approximately 37%, and 9% had COVID-19. This was a high-risk sepsis population spanning ward and intensive care practice.

Programme reach fell at successive stages. Navigators attempted contact with 2,053 intervention patients, reached 1,741 patients or carers, and obtained acceptance from 1,393, or 65.7%. Only 1,016, or 47.9% of the intervention group, received at least one postdischarge follow-up. Among those followed after discharge, the median number of

contacts was ten, the first occurred after two days, and median follow-up duration was sixty-two days. The abstract labels 1,393 as having postdischarge engagement; the main results and supplementary engagement table distinguish acceptance from the smaller number actually followed after discharge.^{2,4}

The primary endpoint occurred in 1,023/2,122 STAR patients, 48.2%, and 684/1,426 controls, 48.0%. The adjusted odds ratio was 1.05, 95% confidence interval 0.90–1.24; $P=0.53$. The observed absolute difference was 0.2 percentage points, with a reported confidence interval of –3.1 to 3.6. The mixed-effects sensitivity analysis gave the same point estimate with a wider interval. The modified analysis among hospital survivors who had not left against advice was also neutral, adjusted odds ratio 1.06, 95% confidence interval 0.91–1.25.^{2,4}

All-cause mortality was 367/2,122, 17.3%, with STAR and 292/1,426, 20.5%, with usual care. The adjusted odds ratio was 0.88, 95% confidence interval 0.77–0.99; $P=0.04$. Readmission occurred in 763/2,122, 35.9%, versus 478/1,426, 33.5%, adjusted odds ratio 1.13, 95% confidence interval 0.92–1.38; $P=0.24$. Thus, the mortality component favoured STAR in an unadjusted-for-multiplicity secondary comparison, whereas readmission was numerically more frequent without a statistically established difference.

No improvement was demonstrated in days alive outside hospital or overall emergency and outpatient utilisation after adjustment. Infection-related readmission was numerically more frequent with STAR, 17.4% versus 15.6%, but its adjusted odds ratio favoured STAR at 0.82, 95% confidence interval 0.69–0.98. The report attributes the direction change particularly to adjustment for implementation period. No convincing interaction identified a benefiting subgroup by the patient or hospital characteristics examined.^{2,4}

Critique

ENCOMPASS tested the practical effect of making STAR available through a health system, including imperfect uptake. That is the appropriate question for an implementation decision. Automated eligibility, broad enrolment, routine data collection and inclusion of community hospitals reduced the distance between the trial and service delivery. Individual consent was not required for research enrolment, limiting the selection that often favours motivated participants in recovery studies. At the same time, patients retained the right to decline the clinical programme.

The design's principal weakness is the small number of randomised units. There were thousands of patients but only seven contributing hospitals, and one hospital had withdrawn after randomisation. Patient numbers do not substitute for independently randomised clusters. Precision depends on the handling of within-hospital correlation and time effects, making the modelling choices consequential. The similar primary point estimates from the marginal and mixed models are reassuring, but the wider mixed-model interval shows that uncertainty is not captured by the point estimate alone.^{2,4}

Time is especially important in a stepped-wedge study: later recruitment is increasingly associated with the intervention. ENCOMPASS spanned major phases of the COVID-19 pandemic and a change in electronic record platform. Calendar-period and COVID-19 adjustment were sensible, but cannot guarantee removal of every change in access, admission thresholds, staffing or background care. Randomising the rollout order protects against some biases; it does not make patients recruited in different periods automatically exchangeable without adequate modelling.

The infection-readmission result illustrates this dependence. Raw proportions suggested more events with STAR, whereas the adjusted model suggested fewer. Such a reversal is possible when time and other covariates are distributed differently; it is not, by itself, proof of a calculation error. It does mean that the adjusted odds ratio must be presented alongside the observed data and understood as a model-based estimate. The main narrative also prints a different upper confidence bound from Table 2; the table's 0.98 is used here. This secondary result should not carry the conclusion of the trial.

The primary result is convincingly neutral for the proposed composite, but the composite itself combines outcomes of very different importance and potentially different mechanisms. Death and a timely, appropriate readmission are not equivalent experiences. A programme designed to detect deterioration might prevent a fatal delay by increasing hospital assessment. Alternatively, the mortality difference may reflect chance, residual time effects or other processes. The study did not randomise readmission, adjudicate its appropriateness as a mediator, or demonstrate that additional admissions caused better survival. The attractive rescue explanation remains a hypothesis.³

The mortality signal deserves follow-up rather than dismissal, but it does not establish a mortality-reducing intervention. It was a secondary comparison among several, its confidence interval only narrowly excluded no effect, and no multiplicity correction was applied. Post hoc time-to-event analysis gave a mortality hazard ratio of 0.90 with an interval extending to 1.00. Concordant direction across analyses is useful, but these

analyses reuse the same patients. They are not independent confirmations. Equally, the neutral composite should not be paraphrased as proof that every aspect of recovery was unaffected.⁴

Inclusion of index-admission deaths adds another interpretative distinction. Although framed as a postsepsis recovery trial, its primary population included patients who never survived to receive outpatient support. STAR began before discharge, so this is not inherently inappropriate: it estimates the policy effect from enrolment. However, the reported 17.3% and 20.5% mortality figures must not be described as death rates solely among hospital survivors. The neutral survivor-only composite sensitivity analysis supports the main conclusion but does not, on its own, establish the separate mortality effect after discharge.

Limited reach is both an explanation and a finding. Fewer than half of intervention-assigned patients received postdischarge follow-up, despite high rates of attempted contact. The patients who did engage received substantial support, but comparing them with nonparticipants would sacrifice randomisation. Survival, illness severity, availability of carers, preferences and ability to answer calls may all influence engagement. An efficacy benefit in a selected adherent subgroup cannot be inferred from the programme's intensity among those reached. For a health system, achieving sustained contact is part of the intervention's effectiveness, not a nuisance to exclude from analysis.

The reporting discrepancy between acceptance and postdischarge follow-up materially affects that assessment. Sixty-six per cent accepted support; approximately 48% received follow-up after discharge. Some accepting patients subsequently died in hospital. Repeating the abstract's larger number as postdischarge reach would overstate implementation. The supplementary table provides the clearer account. The modified-analysis denominator also does not reconstruct completely from the main discharge categories, an additional detail requiring clarification rather than an invented explanation.⁴

Outcome capture has strengths and limits. National death data improve mortality ascertainment beyond the treating hospital. Readmission, however, was defined using Atrium encounters, so admissions elsewhere could be missed. The same data system applied to both groups, but a navigator service could alter where patients seek care. Objective records therefore reduce observer judgement without eliminating every ascertainment concern. Diagnosis codes for cause-specific admissions are less clinically detailed than independent adjudication.

Finally, service benefit cannot be reduced to utilisation alone. The trial did not measure quality of life, symptom burden, caregiver experience or functional recovery, and its main effectiveness report did not establish cost-effectiveness. The required infrastructure included trained navigators, shared records and clinicians able to respond to identified problems. A telephone service without those downstream resources is not the same intervention. ENCOMPASS supports continued evaluation of coordinated recovery care, but does not justify promising fewer combined deaths and readmissions from this particular rollout.

Body of Evidence

IMPACTS is the closest predecessor. Its indexed report describes 691 randomised patients across three North Carolina hospitals, comparing usual care with thirty days of STAR support. Death or readmission occurred in 28.7% versus 33.3%. The unadjusted odds ratio was 0.80, 95% confidence interval 0.58–1.11; the covariate-adjusted estimate was also 0.80, with a narrower interval of 0.64–0.98. Thus, its positive interpretation depended on the adjusted analysis. The ENCOMPASS account gives slightly different summary numbers for IMPACTS; the original report's indexed results are used here.¹

ENCOMPASS expanded hospital diversity, duration of support and sample size, while changing from individual randomisation to randomised rollout. Its neutral primary result does not identify which change mattered. The earlier hospitals' access to outpatient services, programme reach and the pandemic context are plausible contributors, but cross-trial comparisons cannot isolate their effects. Extending a promising intervention can reveal constraints that were less visible in its initial setting.

SMOOTH examined another sepsis-specific approach: primary-care training, nurse case management and physician decision support over twelve months. Among 291 survivors recruited from nine German intensive care units, the intervention did not significantly improve the primary mental health-related quality-of-life outcome at six months. The estimated between-group effect was 2.15 points, 95% confidence interval –1.79 to 6.09. Its endpoint and population differ from ENCOMPASS, but it demonstrates that providing a structured recovery service does not automatically improve the outcome it targets.⁵

The UK PRaCTICaL trial randomised 286 intensive care survivors to nurse-led follow-up or standard care. At twelve months, neither the physical nor mental component of the SF-36 showed a demonstrated benefit, and the programme was more costly. These findings concern a different follow-up model and cannot invalidate every recovery service. They do reinforce the need to test both patient benefit and resource use rather than infer value from the existence of unmet need.⁶

More recent studies remain mixed. The mobile critical care recovery programme enrolled 466 survivors of acute respiratory failure and compared twelve months of nurse-led interdisciplinary support with a telephone-based control. It did not improve the primary quality-of-life outcomes. Hospitalisation was more frequent, 50.2% versus 40.8%, while mortality was numerically lower. That pattern resembles the tension in ENCOMPASS, but neither trial establishes hospitalisation as the mechanism of a survival benefit.⁷

Suivi-Rea randomised 540 intensive care survivors to multidisciplinary consultations or standard follow-up. The composite of death or severe impairment in at least one quality-of-life domain was more frequent with the intervention, adjusted odds ratio 1.49, 95% confidence interval 1.04–2.13. This result cautions against assuming that more professional contact is invariably beneficial; programme content, timing, burden and patient selection matter. It does not imply that STAR and face-to-face multidisciplinary clinics have identical effects.⁸

Outside sepsis, TARGET-READ enrolled 1,386 high-risk medical patients in four Swiss hospitals. Medication reconciliation, education, planned primary-care follow-up and telephone calls did not reduce thirty-day unplanned readmission or death: 21% versus 19%, risk difference 1.7 percentage points, 95% confidence interval –2.5 to 5.9. Together, these trials show how difficult it is to convert sensible transitional-care components into reliable outcome improvements. ENCOMPASS contributes a large pragmatic evaluation and a reason to separate survival, appropriate hospital use and patient-reported recovery in future work.⁹

Does a ninety-day nurse-led STAR programme reduce death or readmission after a sepsis admission?

ENCOMPASS did not demonstrate a reduction in the combined outcome. Lower mortality in a secondary analysis is encouraging but requires confirmation, and the trial did not prove that additional readmissions prevented deaths.

References

1. Taylor SP, Murphy S, Rios A, McWilliams A, McCurdy L, Chou SH, et al. Effect of a Multicomponent Sepsis Transition and Recovery Program on Mortality and Readmissions After Sepsis: The Improving Morbidity During Post-Acute Care Transitions for Sepsis Randomized Clinical Trial. *Crit Care Med.* 2022;50(3):469-479.

2. Taylor SP, Eaton T, Rios A, Boyd D, Tapp H, McWilliams A, et al. Proactive Telehealth-Based Sepsis Transition and Recovery Support, Hospital Readmission, and Mortality: A Randomized Clinical Trial. *JAMA Intern Med.* 2025;185(10):1238-1246.
3. Grady D, Durant RW, Anderson TS. What if Readmission Prevents Death?-The Challenge of Composite Outcomes. *JAMA Intern Med.* 2025;185(10):1246-1247.
4. Taylor SP, Eaton T, Rios A, Boyd D, Tapp H, McWilliams A, et al. Supplementary methods and results to Proactive Telehealth-Based Sepsis Transition and Recovery Support, Hospital Readmission, and Mortality. *JAMA Intern Med.* 2025;185(10):1238-1246.
5. Schmidt K, Worrack S, Von Korff M, Davydow D, Brunkhorst F, Ehlert U, et al. Effect of a Primary Care Management Intervention on Mental Health-Related Quality of Life Among Survivors of Sepsis: A Randomized Clinical Trial. *JAMA.* 2016;315(24):2703-2711.
6. Cuthbertson BH, Rattray J, Campbell MK, Gager M, Roughton S, Smith A, et al. The PRaCTICaL study of nurse led, intensive care follow-up programmes for improving long term outcomes from critical illness: a pragmatic randomised controlled trial. *BMJ.* 2009;339:b3723.
7. Khan BA, Perkins AJ, Khan SH, Unverzagt FW, Lasiter S, Gao S, et al. Mobile Critical Care Recovery Program for Survivors of Acute Respiratory Failure: A Randomized Clinical Trial. *JAMA Netw Open.* 2024;7(1):e2353158.
8. Sharshar T, Grimaldi-Bensouda L, Siami S, Cariou A, Salah AB, Kalfon P, et al. A randomized clinical trial to evaluate the effect of post-intensive care multidisciplinary consultations on mortality and the quality of life at 1 year. *Intensive Care Med.* 2024;50(5):665-677.
9. Donzé J, John G, Genné D, Mancinetti M, Gouveia A, Méan M, et al. Effects of a Multimodal Transitional Care Intervention in Patients at High Risk of Readmission: The TARGET-READ Randomized Clinical Trial. *JAMA Intern Med.* 2023;183(7):658-668.

INHALE-WP3

Enne VI, Stirling S, Barber JA, High J, Russell C, Brealey D, et al. INHALE WP3, a multicentre, open-label, pragmatic randomised controlled trial assessing the impact of rapid, ICU-based, syndromic PCR, versus standard-of-care on antibiotic stewardship and clinical outcomes in hospital-acquired and ventilator-associated pneumonia. *Intensive Care Med.* 2025;51(2):272–286.

Introduction

Hospital-acquired pneumonia creates competing demands for intensive care clinicians. Initial antibiotics must cover a potentially lethal infection before its cause is known, yet unnecessarily broad treatment exposes patients and their microbial ecology to avoidable harm. Conventional cultures may take several days to provide an actionable answer, particularly when susceptibility testing is required. Many patients have already received antibiotics, further complicating interpretation of a negative result. Rapid identification of respiratory pathogens and resistance genes offers an attractive route to earlier treatment refinement. Whether that refinement benefits patients remains a separate question.¹

Syndromic polymerase chain reaction (PCR) panels can identify multiple organisms from a lower respiratory tract sample within hours. However, detecting genetic material is not synonymous with identifying the cause of pneumonia. Colonisation, previous treatment, multiple organisms and pathogens outside the panel can all complicate a treatment decision. The clinical intervention therefore comprises a test, its interpretation and the prescribing response. An analytically successful assay may have little effect if clinicians do not act on its results, or an undesirable effect if an apparently reassuring result prompts inappropriate narrowing.

Earlier randomised trials suggested that molecular testing could shorten inappropriate treatment and accelerate antibiotic changes. These studies were heterogeneous in their populations, sampling strategies and accompanying prescribing support. Several primarily established changes in prescribing rather than precise estimates of clinical safety.^{2,3} INHALE work package 3 addressed both sides of the question: could bedside molecular testing improve antibiotic stewardship in suspected hospital-acquired or ventilator-associated pneumonia while preserving clinical cure? A subsequent economic evaluation added information about resource use, but its interpretation depends on the clinical result.^{1,4}

Synopsis

INHALE-WP3 was an open-label, pragmatic, individually randomised trial in 14 intensive care units at 13 hospitals in England. Eleven units treated adults and three treated children. Recruitment ran from July 2019 to August 2021, with a pause during the first wave of the COVID-19 pandemic. Patients had clinically suspected hospital-acquired or ventilator-associated pneumonia and were about to start empirical antibiotics or change treatment because their condition had deteriorated. A suitable lower respiratory tract specimen was required. Patients expected to die within 48 hours, those receiving care with treatment limitations specified by the trial, and previously enrolled patients were among those excluded.^{1,5}

A central web-based system allocated patients equally, using variable blocks stratified by site. The intervention added the BioFire FilmArray Pneumonia Panel Plus to usual microbiological investigation. Trained staff ran the assay on an intensive care unit instrument and communicated results promptly. A prescribing algorithm, adapted to local practice, supported decisions, but following it was not mandatory. Clinicians could start empirical treatment immediately when required. Both groups underwent conventional culture and susceptibility testing. Control samples also underwent PCR retrospectively, without results being made available to treating clinicians.

The trial had two co-primary outcomes. The first tested superiority for appropriate and proportionate antibiotic treatment 24 hours after randomisation. Treatment had to provide relevant antimicrobial activity without unnecessarily broad coverage. An independent stewardship committee, masked to allocation and clinical outcome, adjudicated this endpoint using the available microbiology, including retrospective PCR in controls. The second tested non-inferiority for clinical cure at day 14, with an absolute margin of 13 percentage points. Cure incorporated the absence of pneumonia-attributable or contributory death, respiratory septic shock, relapse or deterioration and other evidence of unresolved pneumonia. Local investigators assessed cure without masking; it was not equivalent to survival free of every complication.

The target sample was 552. During the pandemic, a blinded assessment led to a reduction in the assumed cure rate from 70% to 55%. The reported analysis used two-sided 95% confidence intervals. Binomial mixed models estimated absolute differences, accounting for site; separate logistic models estimated odds ratios. The clinical non-inferiority decision depended on the confidence interval for the absolute difference. Analyses mainly used participants with available outcome data, with additional adjustment and imputation analyses. The per-protocol clinical analysis excluded intervention patients whose PCR was not completed within 24 hours, rather than excluding all patients whose prescriptions departed from the algorithm.

The report describes 554 eligible randomised patients and 545 retained after consent withdrawals: 276 in the intervention group and 269 controls. There were 453 adults and 92 children; adult median age was 61 years, whereas the paediatric population was predominantly infants. Approximately 69% had ventilator-associated pneumonia, 85% were invasively ventilated, and one-third had COVID-19. More than 90% had received antibiotics during the preceding week. Multidrug resistance was relatively uncommon. These features describe a heavily pretreated intensive care population rather than newly presenting, microbiologically untreated pneumonia.

Median time to a PCR result was 1.5 hours, compared with 73.7 hours for conventional culture. Appropriate and proportionate treatment at 24 hours occurred in 205 of 268 intervention patients (76.5%) and 147 of 263 controls (55.9%). The site-adjusted absolute difference was 21 percentage points (95% confidence interval 13 to 28), with an odds ratio of 2.57 (1.77–3.73). The advantage persisted at 72 hours. Appropriate treatment alone also improved, suggesting that the combined endpoint reflected better matching of coverage as well as decisions about spectrum.¹

This improvement did not translate into a clear reduction in overall antibiotic exposure. Narrow-spectrum treatment at 24 hours was used in 17.3% versus 16.5%, and the difference at 72 hours was also uncertain. Mean defined daily doses per intensive care day were 1.2 versus 1.3, with an estimated difference of –0.08 (95% confidence interval –0.26 to 0.11). Some individual antibiotics were used differently, but the principal finding was improved adjudicated appropriateness, not a demonstrated large reduction in total antibiotic consumption.

Clinical cure occurred in 152 of 268 intervention patients (56.7%) and 171 of 265 controls (64.5%). The adjusted absolute difference was –6 percentage points (95% confidence interval –15 to 2). Because the lower confidence limit crossed the –13-point margin, clinical non-inferiority was not established. The per-protocol analysis gave the same substantive result. A more extensively adjusted sensitivity analysis produced a confidence interval within the margin, but did not replace the unsuccessful primary analysis. The separately modelled odds ratio for cure was 0.68 (0.47–0.98); interpretation of the co-primary non-inferiority test must remain anchored to its specified absolute scale.

By day 28, 85 of 272 intervention patients (31.3%) and 75 of 266 controls (28.2%) had died. The mortality hazard ratio was 1.18 (95% confidence interval 0.87–1.61). There was no convincing improvement in intensive care length of stay or ventilator-free days. Antibiotic-associated diarrhoea was recorded in 26 of 263 versus 14 of 257 patients, although this secondary comparison was not protected against multiple testing. No

serious adverse events were classified as related to the trial intervention; adverse-event ascertainment nevertheless concentrated on specified and unusual events.

The economic evaluation adopted an NHS secondary-care perspective and valued resources using 2020/21 prices. The base case included intensive care costs from randomisation to intensive care exit and the intervention test. Mean costs were £33,149 with PCR and £40,951 with usual care, a difference of –£7,802 (95% confidence interval –£15,696 to £92). Site-adjusted analyses suggested lower costs alongside better stewardship, but lower costs alongside fewer clinical cures. The uncertainty and direction of the clinical effect prevent a simple conclusion that the strategy was unambiguously cost-effective.^{4,6}

Critique

INHALE has an unusually valuable design for a diagnostic stewardship trial because it required clinical reassurance as well as improvement in a prescribing endpoint. Rapid information can change prescriptions almost by definition; demonstrating that these changes preserve patient outcomes is harder. The investigators addressed this explicitly, and the negative non-inferiority result is therefore central rather than an inconvenient secondary observation. The trial supports a process benefit while leaving the intended clinical reassurance unresolved.

Randomisation was concealed before allocation, and the multicentre design included ordinary intensive care workflows. Outcome assessment had an important asymmetry. The stewardship committee was masked and had microbiological information from both groups, reducing the risk that the endpoint simply rewarded knowledge unavailable in controls at adjudication. Conversely, bedside clinicians and those assessing cure knew allocation. Cure included judgement about whether pneumonia had resolved and whether pneumonia contributed to deterioration or death. Differential judgement could affect the result, but its direction cannot be established from the report. It would be equally unjustified to dismiss the lower cure rate as bias or to treat it as definitive proof that PCR-guided prescribing caused harm.

The non-inferiority margin was generous. A strategy capable of reducing cure by as much as 13 percentage points would require a substantial compensating advantage before widespread acceptance. Even against this margin, the primary analysis failed. The interval includes no difference but also clinically important harm. Failure to demonstrate non-inferiority is not proof of inferiority; equally, a non-significant risk-difference comparison is not evidence of equivalence. The distinction between the absolute-difference model and the logistic model matters because their confidence

intervals answer different statistical questions. Selecting the more reassuring adjusted sensitivity analysis would weaken the protection offered by specifying a primary analysis in advance.¹

Intervention delivery also needs careful definition. The assay was delivered rapidly, but the recommended prescribing behaviour was optional. In a supplementary assessment, strict algorithm concordance was identified in 58 of 190 intervention patients with relevant microbiological findings, or 30.5%. This assessment incorporated detailed microbiological information and is not identical to the primary stewardship endpoint. The much larger improvement in appropriate and proportionate treatment shows that useful prescribing changes could occur without exact algorithm concordance. It does not establish that the tested pathway was consistently implemented as its designers envisaged.⁵

The per-protocol population offers limited reassurance about this behavioural issue because its exclusions principally concerned timely testing. Removing two delayed tests does not create a population that reliably followed the prescribing algorithm. Conversely, comparing outcomes among adherent and non-adherent patients would not recover a randomised estimate of algorithm efficacy. Adherence may depend on clinical trajectory, organism complexity, other infections and clinicians' confidence. Post-randomisation analyses relating stewardship success to cure have the same limitation. They may generate explanations for the unexpected clinical pattern, but cannot identify its cause.

The assay itself is one component of a diagnostic judgement. The supplementary microbiology identified organisms outside the panel, and some on-panel organisms were detected by usual testing but missed by PCR. Detection of DNA also cannot reliably distinguish infection from colonisation or remnants of previously treated infection. This matters in a population with extensive prior antibiotic exposure. A panel-negative result is not synonymous with absence of bacterial pneumonia, and detection of a resistance gene or its absence cannot replace every aspect of phenotypic susceptibility testing. The trial continued conventional investigation in both groups, an essential qualification when considering implementation.

The apparent stewardship gain should not be translated into an unmeasured ecological benefit. Total antibiotic exposure and the frequency of narrow-spectrum treatment were not clearly reduced, and long-term resistance outcomes were not established. More appropriate initial activity may be valuable even without less prescribing, but the endpoint combines several desirable actions whose clinical consequences can differ. Earlier escalation for an uncovered pathogen and earlier de-escalation of unnecessary

coverage should not be assumed to have identical benefit–risk profiles. The reported composite does not identify which specific prescribing decisions explain the cure pattern.

Generalisability is affected by both the setting and the pandemic. COVID-19 accounted for one-third of enrolment, introducing competing reasons for respiratory deterioration and death. Prespecified subgroup analyses did not establish a reliably different clinical effect that would justify selecting a subgroup for routine adoption. The paediatric sample was particularly small for non-inferiority, with wide confidence intervals. England's relatively low resistance prevalence also limits extrapolation to units where empirical treatment is frequently inactive against resistant pathogens. Greater potential value in such settings is plausible, but was not demonstrated by this trial.

The economic analysis provides useful resource information without resolving these issues. Test costs included equipment assumptions even though instruments and panels were supplied to the trial; costs should not be interpreted as current procurement prices. The mean saving was heavily influenced by the distribution of intensive care stays. Median stays were similar, and censoring costs at 14 days substantially reduced the estimated saving. Mortality can also shorten resource use. Descriptive cost comparisons stratified by survival do not establish that savings were causally independent of death, because survival is itself a post-randomisation outcome.^{4,6}

Moreover, the economic evaluation used different available-case populations for stewardship and cure, did not estimate quality-adjusted life-years, and could not establish a complete long-term balance of health gains, harms and expenditure. A lower-cost strategy with a less favourable clinical outcome lies in a different decision space from one that is both cheaper and clinically better. The uncertain cost difference, incomplete cost data and unsuccessful clinical non-inferiority test should remain visible together. The evaluation is evidence about possible resource consequences, not a reason to discount the clinical uncertainty.

The study was funded through the National Institute for Health and Care Research. The manufacturer supplied instruments and testing materials without a reported role in design, analysis or writing. Relevant author relationships were disclosed. Public funding and independent analysis strengthen credibility, while the pragmatic, industry-supported provision of equipment still leaves local implementation costs, staffing and prescribing support to be assessed. The transferable lesson is that rapid diagnostics require evaluation as an entire care pathway, with patient outcomes monitored alongside turnaround time and prescription quality.

Body of Evidence

FLAGSHIP II examined a different but related pathway: multiplex bacterial PCR on bronchoalveolar lavage samples in adults with pneumonia at risk of Gram-negative infection. In two Swiss centres, 208 patients were randomised to PCR-supported management or usual investigation. The intervention included an antibiotic recommendation after results became available. Mean duration of inappropriate treatment was 47.1 versus 85.7 hours, an estimated reduction of 38.6 hours. Deaths were uncommon relative to the sample size, and the trial was not a precise test of mortality. Its contribution is evidence that rapid microbiology linked to prescribing advice can reduce time on inappropriate therapy.²

Poole and colleagues randomised 200 critically ill adults with community-acquired, hospital-acquired or ventilator-associated pneumonia to molecular point-of-care testing or routine care. Results-directed treatment occurred much more frequently with rapid testing, and the time to such treatment was shortened substantially. Antibiotic de-escalation was also more frequent. Major clinical outcomes did not differ clearly, but a trial of this size cannot exclude modest clinical harm. Its mixed pneumonia population and prescribing support differ from INHALE, preventing a direct assumption that the same clinical effect should follow every use of the technology.³

The MultiCoV trial tested respiratory multiplex PCR together with repeated procalcitonin measurements in severe SARS-CoV-2 pneumonia. Of 194 randomised patients, 191 were included in its intention-to-treat analysis. Although respiratory bacterial co-infection was identified more frequently with the intervention, antibiotic-free days at day 28 did not improve: medians were 12 versus 14 days, with a wide confidence interval for the difference. Mortality and intensive care length of stay also did not differ clearly. This illustrates the gap between additional diagnostic information and a measurable reduction in antibiotic exposure, particularly in a population with frequent subsequent infections.⁷

In a Norwegian emergency department, Markussen and colleagues studied 374 adults with suspected community-acquired pneumonia. Syndromic PCR increased pathogen-directed treatment within 48 hours from 13.4% to 35.3% and shortened the time to such treatment. This supports a prescribing effect in a different clinical setting. It is indirect evidence for patients who develop pneumonia during prolonged hospitalisation or mechanical ventilation, whose microbiology, previous treatment and competing risks are substantially different.⁸

Across these trials, the most consistent benefit is earlier microbiologically informed prescribing. The size and nature of that benefit depend on patient selection, respiratory

sampling, the assay and the support provided to clinicians. Neither increased diagnostic yield nor a more favourable stewardship score is a reliable substitute for a patient-centred endpoint. Smaller trials with no detectable clinical difference should not be interpreted as having established safety simply because their process outcomes were positive.

INHALE therefore advances the evidence in an uncomfortable but useful direction. It confirms that a rapid intensive care testing pathway can improve antibiotic matching in suspected hospital-acquired and ventilator-associated pneumonia. It also shows why clinical non-inferiority must be tested rather than presumed. By the end of 2025, the evidence supported further refinement and evaluation of diagnostic stewardship pathways, with explicit attention to clinical decision support and safety. It did not establish that this particular pragmatic strategy could routinely replace usual diagnostic management without risking a clinically important reduction in cure.

Should rapid syndromic PCR routinely guide antibiotics for suspected hospital-acquired or ventilator-associated pneumonia in intensive care?

INHALE improved early antibiotic appropriateness but did not establish non-inferiority for clinical cure. Its possible cost savings do not justify routine adoption without a carefully evaluated prescribing pathway and clinical safety monitoring.

References

1. Enne VI, Stirling S, Barber JA, High J, Russell C, Brealey D, et al. INHALE WP3, a multicentre, open-label, pragmatic randomised controlled trial assessing the impact of rapid, ICU-based, syndromic PCR, versus standard-of-care on antibiotic stewardship and clinical outcomes in hospital-acquired and ventilator-associated pneumonia. *Intensive Care Med.* 2025;51(2):272-286.
2. Darie AM, Khanna N, Jahn K, Osthoff M, Bassetti S, Osthoff M, et al. Fast multiplex bacterial PCR of bronchoalveolar lavage for antibiotic stewardship in hospitalised patients with pneumonia at risk of Gram-negative bacterial infection (Flagship II): a multicentre, randomised controlled trial. *Lancet Respir Med.* 2022;10(9):877-887.
3. Poole S, Tanner AR, Naidu VV, Borca F, Phan H, Saeed K, et al. Molecular point-of-care testing for lower respiratory tract pathogens improves safe antibiotic de-escalation in patients with pneumonia in the ICU: Results of a randomised controlled trial. *J Infect.* 2022;85(6):625-633.

4. Wagner AP, Enne VI, Gant V, Stirling S, Barber JA, Livermore DM, et al. Cost-effectiveness of rapid, ICU-based, syndromic PCR in hospital-acquired pneumonia: analysis of the INHALE WP3 multi-centre RCT. *Crit Care*. 2025;29(1):352.
5. Enne VI, et al. Supplementary materials and accompanying respiratory microbiology workbook to INHALE WP3. *Intensive Care Med*. 2025.
6. Wagner AP, et al. Supplementary materials to the INHALE WP3 economic evaluation. *Crit Care*. 2025.
7. Fartoukh M, Nseir S, Mégarbane B, Cohen Y, Lafarge A, Contou D, et al. Respiratory multiplex PCR and procalcitonin to reduce antibiotic exposure in severe SARS-CoV-2 pneumonia: a multicentre randomized controlled trial. *Clin Microbiol Infect*. 2023;29(6):734-743.
8. Markussen DL, Serigstad S, Ritz C, Knoop ST, Ebbesen MH, Faurholt-Jepsen D, et al. Diagnostic Stewardship in Community-Acquired Pneumonia With Syndromic Molecular Testing: A Randomized Clinical Trial. *JAMA Netw Open*. 2024;7(3):e240830.



Paediatric Trials

GASTROSAM

Maitland K, Ouattara SM, Sainna H, Chara A, Ogundipe OF, Sunyoto T, et al. Intravenous Rehydration for Severe Acute Malnutrition with Gastroenteritis. *N Engl J Med*. 2025;393(13):1257-1268.

Introduction

Fluid treatment in a severely malnourished child with diarrhoea presents an uncomfortable balance. Inadequate replacement may leave dangerous hypovolaemia untreated, while concerns about myocardial function and sodium handling have encouraged caution with intravenous fluid. The guidance underlying GASTROSAM favoured oral or nasogastric rehydration, reserving intravenous boluses for shock. This differs from the intravenous rehydration commonly used for severe dehydration in children without severe acute malnutrition.¹

The distinction has major practical consequences. A child who cannot drink may require repeated administration through a nasogastric tube, assessment of tube position and close observation for vomiting or deterioration. These tasks can be difficult in crowded wards with limited staffing. Conversely, intravenous treatment requires reliable access, accurate fluid delivery and monitoring. Neither route becomes intrinsically safe merely because it is familiar or recommended.

The concern that malnourished children routinely develop cardiac failure after intravenous fluid has limited direct support. Small physiological studies found hypovolaemia and preserved global myocardial function in many such children, without the expected frequency of pulmonary oedema. However, physiological reassurance is not equivalent to evidence of improved survival. The FEAST trial also demonstrated that bolus fluids can cause harm in a different population of African children, making careful attention to patient selection essential.^{2,3,4}

GASTROSAM compared standard oral rehydration with two intravenous strategies in children with severe acute malnutrition and gastroenteritis-associated dehydration. Its most useful contribution is evidence about a long-standing safety assumption. Its ability to determine mortality benefit, equivalence between routes or the best infusion rate is more limited.¹

Synopsis

GASTROSAM was an investigator-initiated, open-label, factorial randomised trial at six hospitals in Kenya, Uganda, Niger and Nigeria. Recruitment occurred between September 2019 and October 2024, with a prolonged interruption during the COVID-19 pandemic. More than 90% of participants came from Niger and Nigeria. Funding came from the Joint Global Health Trials scheme, the UK Medical Research Council and Médecins Sans Frontières. An independent monitoring committee reviewed interim data. Parents or guardians provided written consent; urgent enrolment could use verbal assent followed by deferred written consent under approved arrangements.¹

Eligible children were aged six months to twelve years and had severe acute malnutrition, gastroenteritis and clinical signs of severe dehydration. Malnutrition was identified through severe anthropometric deficits or bilateral nutritional oedema. Gastroenteritis required more than three loose stools daily. Severe dehydration required at least two of altered consciousness, sunken eyes, markedly reduced skin turgor or inability to take or retain oral fluids. Known congenital or rheumatic heart disease and diarrhoea lasting more than fourteen days were exclusions.

Randomisation used computer-generated lists with variable blocks and sequentially numbered opaque envelopes. Allocation was in a 2:1:1 ratio to oral rehydration, rapid intravenous rehydration or slow intravenous rehydration. The two intravenous groups were pooled for the main comparison, giving approximately equal numbers in the pooled intervention and control groups. Clinicians and families knew the strategy; laboratory testing was masked.^{1,5}

The oral strategy used frequent small volumes of oral rehydration solution, administered by nasogastric tube when necessary, with nutritional milk according to the treatment schedule. Children with shock could receive an intravenous bolus of 15 mL/kg over an hour. The rapid intravenous strategy used lactated Ringer's solution, 100 mL/kg over three to six hours according to age, following the general structure of WHO plan C. Boluses of 20 mL/kg were permitted for shock. The slow strategy used the same total rehydration volume over eight hours, without boluses, including for children meeting the trial's shock definition.

Shock required cold extremities, a weak fast pulse and capillary refill exceeding three seconds. This was a specific bedside definition; a child with an isolated low blood pressure did not necessarily fulfil it. Intravenous-group children subsequently received oral replacement for ongoing losses and nutritional care. The trial therefore compared initial rehydration strategies rather than exclusive intravenous feeding with exclusive oral treatment.

A separate factorial randomisation assigned the type of oral solution: ReSoMal, formulated for malnutrition, or standard WHO oral rehydration solution. Both the initial oral group and children subsequently receiving oral fluid after intravenous rehydration participated in that comparison. Its principal results were not presented in this report. The intravenous-versus-oral findings cannot identify the preferred oral formulation.^{1,5}

All groups received enhanced observation. Bedside assessments occurred every thirty minutes for two hours, hourly until eight hours, and at specified later time points. Glucose, haemoglobin and lactate could be checked at the bedside; biochemical assessments occurred at baseline, eight hours and twenty-four hours. In Niger and Nigeria, dedicated teams treated participants in closely monitored units, although assisted ventilation was unavailable. Other sites used general paediatric wards. Staff actively assessed for suspected pulmonary oedema, heart failure and other serious adverse events.

The primary endpoint was death by ninety-six hours. Secondary endpoints included death by twenty-eight days, anthropometric changes and urine output. Safety assessments included pulmonary or cardiac overload, sodium changes and correction of severe hyponatraemia or hypokalaemia. The final sample-size calculation assumed 58% mortality with oral treatment and 41% with pooled intravenous treatment. A sample of 272 was intended to provide 80% power for that large difference. The accompanying editorial describes the study's evolution from an initial physiological phase II design to a mortality superiority comparison.^{1,6}

The primary analysis used a site-adjusted Mantel–Haenszel risk ratio; twenty-eight-day mortality used Cox regression. Analyses followed allocated treatment. Correction of electrolyte abnormalities was analysed with death as a competing risk. Secondary confidence intervals were not adjusted for multiplicity. Complete-case analyses were supplemented by multiple imputation for selected biochemical and anthropometric outcomes.^{1,5}

The trial enrolled 272 children: 138 oral, 67 rapid intravenous and 67 slow intravenous. Median age was thirteen months and typical weight was just over 5 kg. Approximately 79% could not take or retain oral fluids, 38% had altered consciousness, and 25 children met the shock definition at randomisation. Severe hyponatraemia was present in 137/262 children with measurements and severe hypokalaemia in 115/258. Only eleven children had kwashiorkor and two had known HIV infection.¹

Initial treatment separation was substantial. Intravenous rehydration began in 66/67 rapid-group children and all 67 slow-group children; one rapid-group child died before

treatment. In the oral group, 31/138 received some intravenous fluid within twenty-four hours, including treatment for initial or subsequent shock and small volumes for other indications. This permitted rescue was part of the strategy. Nasogastric rehydration was used in 126/135 oral-group recipients, 93%, compared with 82/126 intravenous-group children with relevant data, 65%. Vomiting was also less frequent in the intravenous groups.^{1,5}

Vital status was known for 271 children at ninety-six hours and 267 at twenty-eight days. The supplementary flow diagram states that the one child with unknown early status was assumed alive for the primary risk-ratio estimate. Eleven oral-group children and nine intravenous-group children died by ninety-six hours, approximately 8% and 7%. The published site-adjusted risk ratio was 1.02, 95% confidence interval 0.41–2.52, with a reported P value of 0.69. Five deaths occurred in the rapid group and four in the slow group. At twenty-eight days, deaths numbered seventeen and fourteen, hazard ratio 0.85, 95% confidence interval 0.41–1.78.^{1,5}

No suspected pulmonary oedema or signs consistent with heart failure were recorded. Serious adverse events affected 32/138 oral-group children, 14/67 rapid-group children and 10/67 slow-group children. At eight hours, severe hyponatraemia was less frequent with intravenous treatment, 20/128 versus 58/129, but severe hypokalaemia was more frequent, 57/127 versus 40/128. Potassium increased more slowly in the intravenous groups. These biochemical findings are secondary and should be considered together; the electrolyte effects were not uniformly favourable.¹

Critique

Randomisation directly addressed a treatment distinction that had persisted despite weak comparative evidence. The trial enrolled children with severe nutritional deficits, frequent electrolyte abnormalities and difficulty tolerating oral fluids, rather than a low-risk convenience sample. Mortality was an objective endpoint, early follow-up was almost complete, and actual fluid delivery differed meaningfully between groups. These features make GASTROSAM more informative than an uncontrolled account of successful intravenous rehydration.

Its dominant limitation is imprecision. The assumed oral-group mortality of 58% was far above the observed ninety-six-hour mortality of 8%. Only twenty primary events occurred. Consequently, the confidence interval around the adjusted risk ratio permits a substantial reduction in mortality as well as more than a doubling. Failure to demonstrate superiority does not establish equivalence or noninferiority. The trial was not designed with an acceptable excess-risk margin for either claim.

The point estimate also requires careful presentation. The crude proportion of deaths was lower in the pooled intravenous groups, whereas the site-adjusted risk ratio was slightly above one. Adjustment can change an estimate when event patterns differ by site, particularly with sparse data. It should not be replaced by a crude ratio or interpreted as evidence of harm. The report's P value is retained as published, without attempting an unsupported reconstruction of the exact test from the rounded confidence interval.

The absence of observed cardiac failure is nevertheless meaningful. If intravenous rehydration commonly caused overt, immediate overload in this population, a closely monitored study might reasonably have detected a signal. Zero events among approximately 134 intravenous-assigned children weakens the proposition that such complications are frequent or inevitable. It does not prove zero risk. The accompanying editorial uses an approximate upper risk bound of 2.3%, illustrating that an uncommon adverse event remains compatible with the data. Moreover, absence of overt cardiac failure does not exclude other mechanisms of harm.⁶

Monitoring is central to both the trial's credibility and its applicability. Frequent observations, laboratory testing and dedicated staff improved detection and treatment of deterioration in every group. These measures could plausibly explain some of the difference between observed mortality and earlier reports, but they were not randomised. The trial cannot quantify how much enhanced care improved survival. Nor can outcomes achieved in a staffed research unit be assumed to persist when an identical fluid schedule is used in an overcrowded ward with infrequent observation.

The comparator should therefore be described as the recommended oral strategy delivered with trial support, rather than unsupported routine care. Intravenous rescue for shock did not represent simple protocol failure; it was an intended component of the control policy. Conversely, the intravenous groups commonly required subsequent oral or nasogastric fluid, so the trial does not show that intravenous treatment eliminates the need for enteral support. It suggests a reduction in nasogastric use, not its replacement in every child.

Pooling the intravenous groups strengthened the principal comparison but limited conclusions about infusion speed. Each intravenous arm contained only 67 children, and the separate mortality intervals were extremely wide. Four versus five early deaths cannot establish that slow infusion is safer than rapid infusion. The slow regimen is operationally simpler because it avoids age-dependent phases and boluses, but simplicity is a practical characteristic, not evidence of superior survival. The study also

contained too few children with shock to settle the best resuscitation strategy for that subgroup.

Electrolyte findings caution against a one-dimensional safety interpretation. Intravenous treatment corrected severe hyponatraemia sooner, while potassium recovery was slower and severe hypokalaemia at eight hours was more common. The sodium increase over the first eight hours was also larger. These observations justify attention to both sodium and potassium; they do not validate an unrestricted correction rate for every child. One neurological serious adverse event was recorded in the slow group, but the data do not establish its mechanism or attribute it to electrolyte correction.^{1,5}

The factorial design is another reason for precision. Oral solution type was independently allocated, and oral fluids continued after initial intravenous treatment. The main comparison averages across those oral assignments. A subgroup analysis did not identify a clear interaction, but its confidence intervals were broad. It would be inappropriate to conclude either that ReSoMal and standard oral solution are equivalent or that the observed electrolyte differences arise exclusively from route rather than the whole assigned strategy.

Generalisability to oedematous malnutrition is particularly uncertain. Only 4% of participants had kwashiorkor. Earlier cardiac studies offer physiological reassurance, but cannot supply the missing randomised safety data for a substantially larger group with oedema. The broad eligibility age range similarly conceals a population dominated by infants and young toddlers. Known structural heart disease and prolonged diarrhoea were excluded. These boundaries matter more than the nominal inclusion of four countries when considering an individual clinical setting.

Hydration can also alter the anthropometric measurements used to diagnose malnutrition. Some children near a threshold may move above it after rehydration, although the very low median weight-for-height scores in GASTROSAM indicate substantial wasting in this cohort. Short-term weight gain therefore reflects fluid restoration as well as nutritional change. It should not be described as evidence of improved nutritional rehabilitation, especially when the small day-three difference was absent by day seven.^{1,7}

Missingness varied markedly by endpoint. Early vital status was nearly complete, but urine output was available in only thirty-six participants and was not actively collected at the large MSF sites. That outcome cannot serve as a robust physiological confirmation for the whole trial. Multiple imputation supported selected biochemical and anthropometric analyses, but does not repair every unmeasured variable or prove a

missing-completely-at-random mechanism. Overall, GASTROSAM supports reassessment of an assumed frequent overload hazard while leaving the mortality comparison and several important subgroups unresolved.

Body of Evidence

Akech and colleagues conducted an earlier phase II randomised trial in Kenyan children with severe malnutrition and shock. Sixty-one children were enrolled, including both severe diarrhoeal dehydration and presumptive septic shock. The principal comparison used small-volume hypotonic fluid or Ringer's lactate. Shock frequently persisted and mortality remained high; no pulmonary oedema or cardiogenic failure was detected. The study challenged confidence in the restrictive regimen, but its size, mixed population and limited fluid volumes did not establish the optimal rehydration strategy. GASTROSAM subsequently tested a more complete intravenous replacement approach.⁸

AFRIM was a prospective physiological study of twenty children with severe malnutrition and hypovolaemic shock, comparing two sequential management strategies rather than randomly allocated groups. Fifteen had kwashiorkor. Haemodynamic findings often suggested severe hypovolaemia, and pulmonary oedema or congestive cardiac failure was not observed. Mortality was nevertheless very high. Its value lies in questioning a proposed mechanism of harm and identifying persistent circulatory abnormalities; the nonrandomised comparison cannot establish a survival advantage for either regimen.³

CAPMAL examined cardiac function in 88 children with severe acute malnutrition and 22 severity-matched controls. It included 52 children with marasmus and 36 with kwashiorkor. Although myocardial mass was initially lower in malnourished children, global functional measurements were generally preserved, and echocardiography did not identify the anticipated pattern of cardiac failure, including among children receiving intravenous fluid. Electrolyte abnormalities and electrical changes were important. These findings support physiological equipoise, while their observational design leaves clinical treatment effects unresolved.²

FEAST provides a different and essential caution. In 3,141 children with severe febrile illness and impaired perfusion in its main stratum, albumin or saline boluses increased forty-eight-hour mortality compared with no bolus: approximately 10.5–10.6% versus 7.3%, with a pooled relative risk of 1.45. Children with malnutrition or gastroenteritis were excluded. FEAST therefore cannot be used as direct evidence against replacing a genuine diarrhoeal fluid deficit in GASTROSAM's population. Equally, GASTROSAM cannot reverse FEAST's findings for children with severe infection without that fluid-loss syndrome.⁴

The earlier GASTRO trial was more directly relevant to infusion rate. It randomised 122 children with severe dehydration from gastroenteritis to plan C or the same total Ringer's lactate volume over eight hours without boluses. Serious adverse events occurred in three versus two children, and no difference was demonstrated in correction of dehydration or time to discharge. The study supported feasibility of slower rehydration but was too small to settle mortality. Retrospective weight changes also suggested that clinical criteria had overclassified severe dehydration in many participants.⁹

A separate repeated-measures study by Mwangome and colleagues assessed 325 Kenyan children with clinical dehydration. After forty-eight hours of treatment, approximately one-fifth initially classified as severely malnourished by weight-for-length or mid-upper-arm circumference crossed above the relevant threshold. Mid-upper-arm circumference was less affected than weight-based classification, but neither was immune. This matters when guidelines make a sharp treatment distinction based on measurements influenced by the acute fluid deficit itself. It supports reassessment after rehydration, not abandonment of nutritional assessment.⁷

GASTROSAM brings these strands into a randomised comparison focused on severe malnutrition and gastroenteritis. It adds reassurance against frequent overt fluid overload under closely monitored conditions, alongside no demonstrated mortality advantage. The next evidentiary step is a larger comparison capable of excluding clinically important harm and examining delivery under realistic staffing conditions, with sufficient representation of oedematous malnutrition and careful electrolyte monitoring.

Does GASTROSAM establish that intravenous rehydration is superior or equally safe to oral rehydration in severely malnourished children with gastroenteritis?

No. It found no mortality difference and no overt cardiac or pulmonary overload, providing useful reassurance under close monitoring, but the wide confidence interval does not establish superiority, equivalence or safety in every subgroup.

References

1. Maitland K, Ouattara SM, Sainna H, Chara A, Ogundipe OF, Sunyoto T, et al. Intravenous Rehydration for Severe Acute Malnutrition with Gastroenteritis. *N Engl J Med*. 2025;393(13):1257-1268.

2. Brent B, Obonyo N, Akech S, Shebbe M, Mpoya A, Mturi N, et al. Assessment of Myocardial Function in Kenyan Children With Severe, Acute Malnutrition: The Cardiac Physiology in Malnutrition (CAPMAL) Study. *JAMA Netw Open*. 2019;2(3):e191054.
3. Obonyo N, Brent B, Olupot-Olupot P, Boele van Hensbroek M, Kuipers I, Wong S, et al. Myocardial and haemodynamic responses to two fluid regimens in African children with severe malnutrition and hypovolaemic shock (AFRIM study). *Crit Care*. 2017;21(1):103.
4. Maitland K, Kiguli S, Opoka RO, Engoru C, Olupot-Olupot P, Akech SO, et al. Mortality after fluid bolus in African children with severe infection. *N Engl J Med*. 2011;364(26):2483-2495.
5. Maitland K, Ouattara SM, Sainna H, Chara A, Ogundipe OF, Sunyoto T, et al. Supplementary methods and results to Intravenous Rehydration for Severe Acute Malnutrition with Gastroenteritis. *N Engl J Med*. 2025;393(13):1257-1268.
6. Peters MJ. Data versus Dogma - Safety of Intravenous Rehydration in Severely Malnourished Children. *N Engl J Med*. 2025;393(13):1336-1337.
7. Mwangome MK, Fegan G, Prentice AM, Berkley JA. Are diagnostic criteria for acute malnutrition affected by hydration status in hospitalized children? A repeated measures study. *Nutr J*. 2011;10:92.
8. Akech SO, Karisa J, Nakamya P, Boga M, Maitland K. Phase II trial of isotonic fluid resuscitation in Kenyan children with severe malnutrition and hypovolaemia. *BMC Pediatr*. 2010;10:71.
9. Houston KA, Gibb J, Olupot-Olupot P, Obonyo N, Mpoya A, Nakuya M, et al. Gastroenteritis aggressive versus slow treatment for rehydration (GASTRO): a phase II rehydration trial for severe dehydration: WHO plan C versus slow rehydration. *BMC Med*. 2019;17(1):122.

IsoCOMFORT

Miatello J, Palacios-Cuesta A, Radell P, Oberthuer A, Playfor S, Amores-Hernández I, et al; IsoCOMFORT Study Group. Inhaled isoflurane for sedation of mechanically ventilated children in intensive care (IsoCOMFORT): a multicentre, randomised, active-control, assessor-masked, non-inferiority phase 3 trial. *Lancet Respir Med*. 2025;13(10):897–910.

Introduction

Sedation during paediatric mechanical ventilation must balance comfort and safety against delayed awakening, drug accumulation, withdrawal and interference with recovery. Children differ considerably in developmental stage, disease, previous exposure and required depth of sedation. A sedative that reliably reaches a behavioural target may still be unsuitable if it prolongs ventilation or causes important adverse effects. The choice of agent therefore cannot be separated from analgesia, repeated assessment and the clinical justification for the target itself.¹

Midazolam remains a familiar comparator in paediatric sedation research, but its limitations create interest in alternatives. Inhaled isoflurane has rapid pulmonary elimination and can be delivered through an anaesthetic-conserving device attached to an intensive care ventilator. Adult randomised evidence has shown that it can maintain sedation and reduce opioid requirements compared with propofol over a relatively short treatment period.² These results provide a rationale for paediatric evaluation without establishing safety across childhood or during prolonged exposure.

The relevant clinical question is broader than whether two drugs achieve similar scale scores. An inhaled technique requires equipment, gas monitoring, staff expertise and scavenging. Its effects must be assessed in the context of the ventilator circuit and the child's respiratory needs. Potential advantages in awakening and opioid consumption must be weighed against haemodynamic events, neurological effects and uncertainty about longer-term outcomes. Regulatory efficacy and selection as the preferred initial sedative are related but distinct decisions.

IsoCOMFORT compared inhaled isoflurane with intravenous midazolam in mechanically ventilated children aged 3–17 years. Its main endpoint concerned time within an individually prescribed sedation range without rescue medication. The trial offers important randomised evidence in a difficult research population, but its reduced

sample, non-inferiority margin and discordance between analysis populations require attention before its findings are extended to routine first-line use.¹

Synopsis

IsoCOMFORT was a multicentre, active-control, assessor-masked phase 3 trial at 19 paediatric intensive care units in Spain, France, Germany and the United Kingdom. Recruitment ran from January 2021 to January 2023. Children aged 3–17 years were eligible when invasive ventilation and sedation were expected to continue for at least 12 hours. Consent was obtained from parents or legal representatives, with participation of the child where appropriate.^{1,3}

Exclusions included contraindications to the study drugs, suspected susceptibility to malignant hyperthermia, uncompensated shock, ongoing seizures requiring acute treatment, suspected raised intracranial pressure or intracranial-pressure monitoring, and circumstances preventing reliable behavioural assessment. Children already continuously sedated for more than 72 hours and those requiring continuous neuromuscular blockade were also outside the intended comparison. These restrictions excluded several situations in which rescue volatile anaesthesia might otherwise be considered.

A central web-response system allocated participants 2:1 to isoflurane or midazolam, using variable blocks and stratification by country, age group and planned versus unplanned ventilation. Recruitment was structured so that at least half the treated children would be younger than eight years. Treating clinicians knew allocation because the inhaled technique required gas monitoring. Separate trained assessors performed masked COMFORT Behaviour scale assessments, with equipment and displays concealed during scoring. Assessors were also unaware of the prescribed sedation target.

After randomisation, the clinical team selected one of three COMFORT-B target ranges: light sedation, scores 17–22; moderate sedation, 11–16; or deep sedation, 6–10. One clinically indicated target change was allowed. Isoflurane was delivered through the Sedaconda ACD-S device and titrated using monitored gas concentrations and clinical assessment. Midazolam was administered intravenously, with continuation or adjustment of an existing infusion where applicable. Treatment was planned for up to 48 hours, with a six-hour allowance around that limit. Opioid analgesia remained at clinician discretion, and specified rescue sedatives could be administered.

The primary endpoint was the percentage of assessed time within the individual COMFORT-B target without rescue sedation. Masked assessments were scheduled every

two hours. Rescue-sedation intervals counted as failure, while intervals with missing behavioural assessments were omitted from the primary denominator. The full analysis set required at least six hours of allocated study treatment and at least three masked assessments. Although analysed according to assigned treatment, it was consequently not an analysis of every randomised child. A per-protocol set applied additional criteria.

The original design sought superiority, with non-inferiority available as a subsequent test, and planned 144 evaluable participants. Recruitment difficulties, including the pandemic and the age distribution of paediatric intensive care patients, prompted a focus on non-inferiority with approximately 90 evaluable children. The published report states that changes preceded database lock. The non-inferiority rule allowed a 15% relative reduction in adequately sedated time compared with the observed midazolam mean. Thus, the absolute margin was calculated from the control result rather than fixed as an unchanging number of percentage points.

The primary mixed-effects analysis included treatment as a fixed effect and country as a random effect. Age group and reason for admission were not included in that model, although they were randomisation strata; a post hoc model incorporated them. Because normality assumptions were not met, bootstrap and permutation analyses were also performed. Sensitivity analyses varied the handling of missing assessments. Secondary endpoints included opioid exposure, rescue medication and time from discontinuing the sedative for extubation to actual extubation. Safety follow-up extended to 30 days.

Ninety-six children were randomised, 63 to isoflurane and 33 to midazolam. Two assigned isoflurane did not receive it, leaving 94 in the safety population. The full analysis set included 59 and 33 children, respectively, and the per-protocol set included 56 and 29. Mean age in the full analysis set was 7.7 years, 38% were female and 70% were postoperative patients. Deep sedation was initially prescribed for 58%, moderate sedation for 38% and light sedation for only 4%. There was no crossover between assigned sedatives.

The least-squares mean percentage of adequately sedated time was 68.94% with isoflurane and 62.37% with midazolam. The difference was 6.57 percentage points (95% confidence interval -8.99 to 22.13). The margin calculated from the midazolam mean was -9.36 percentage points, so the primary analysis met the non-inferiority criterion. The lower confidence limit cleared that margin by only 0.37 percentage points. Superiority was not established.¹

Bootstrap analysis and adjustment for all stratification variables supported non-inferiority. However, the sensitivity analysis using linear interpolation for missing

COMFORT-B assessments did not. Nor did the per-protocol analysis: its estimated difference was 5.34 percentage points (95% confidence interval –10.48 to 21.17), compared with a margin of –9.84. Additional per-protocol sensitivity and non-parametric analyses also failed to establish non-inferiority. The point estimates remained favourable to isoflurane, but the uncertainty extended beyond the permitted loss of efficacy.³

Mean opioid requirement, expressed as intravenous fentanyl equivalents, was lower with isoflurane: adjusted means of 2.1 versus 4.6 micrograms/kg/hour ($P=0.0004$). Rescue sedation was required by 39% versus 52%, but this difference was not statistically convincing, and the number of rescue doses also did not differ clearly. Among the selected children whose study sedative was stopped with the intention of extubation, median time to extubation was 0.75 versus 1.09 hours. This comparison involved 35 and 20 children; another midazolam patient was re-sedated and excluded from the displayed analysis.

Adverse events were reported in 51 of 61 isoflurane recipients (84%) and 21 of 33 midazolam recipients (64%). Serious events affected 19 (31%) and eight (24%), respectively, with none judged related to treatment. Hypotension occurred in 31% versus 15%, and acute kidney injury was recorded in ten isoflurane recipients versus none receiving midazolam. Severe hypotension considered treatment-related occurred in one child per group. Three isoflurane recipients discontinued treatment because of adverse events. Delirium and withdrawal were not clearly reduced. Two midazolam patients died within 30 days, while one isoflurane recipient died after the study window; all deaths were judged unrelated to treatment.

Critique

The trial provides direct evidence that isoflurane can deliver sustained sedation in selected ventilated children using an intensive care ventilator. Central randomisation, multiple countries and masked assessment strengthen the comparison. Masking the behavioural assessor was particularly useful because the primary endpoint could otherwise be influenced by expectations about a new technique. The authors also reported analyses that did not support their principal non-inferiority conclusion, allowing readers to assess its fragility.

That fragility is clinically important. The primary confidence interval only narrowly excluded the margin, and the per-protocol result did not confirm non-inferiority. In non-inferiority trials, both analyses can contribute to interpretation because exclusions, incomplete exposure and departures from assigned treatment may make groups appear more alike. Here, even the full analysis set excluded four children, all allocated to

isoflurane. The effect of such exclusions cannot be assumed to be conservative. Conversely, the per-protocol result does not prove isoflurane inferior: its point estimate still favoured isoflurane, but its confidence interval was too wide to establish the required reassurance.

The margin itself warrants explanation. A 15% relative allowance was translated into different absolute margins according to the observed control mean in each analysis. Consequently, the margin changed alongside the statistical estimate, from -9.36 in the primary analysis to -9.84 in the per-protocol analysis. This is not equivalent to a fixed, independently justified absolute clinical loss chosen before outcomes were observed. It makes transparent presentation of the rule essential, particularly when success depends on a fraction of one percentage point. Agreement with a regulatory development plan does not by itself establish that clinicians or families would accept the permitted reduction in well-controlled sedation.

The reduced recruitment target compounds this uncertainty. The original superiority calculation assumed a substantial benefit of isoflurane, and the subsequent non-inferiority calculation also drew on an expectation that its true performance could be better. If the actual difference is smaller, such assumptions can leave limited power to establish non-inferiority. The reported shift arose from recruitment feasibility and regulatory discussions; the available published account does not justify implying that it followed an unmasked efficacy result. Nevertheless, the final small sample restricts precision for both efficacy and safety.

The primary endpoint measures success against a prescribed target, not whether the target was optimal for recovery. Most children were assigned deep sedation, and only four initially had a light target. Moreover, targets were selected after randomisation by clinicians who knew the drug allocation. Masked scoring reduces observer bias, but cannot remove any influence of treatment knowledge on target choice, opioid dosing, rescue decisions or readiness to stop sedation. Similar distributions of prescribed targets are reassuring but do not turn the study into a comparison of otherwise identical management pathways.

COMFORT-B is a useful behavioural tool, but measurements every two hours cannot directly observe the entire intervening period. The calculated time in range depends on how observations, rescue doses and missing intervals are translated into time. The primary method removed missing intervals, whereas one interpolation analysis failed to support non-inferiority. With the confidence limit so close to the margin, small changes in these assumptions can alter the formal conclusion. This is a reason to retain the

uncertainty, rather than infer that several positive sensitivity analyses outvote an inconvenient negative one.

The opioid reduction is a potentially valuable finding. Lower exposure could plausibly reduce accumulation, constipation or withdrawal, but those downstream benefits were not established here. Analgesia was adjusted by unmasked clinicians, and opioids may be increased for agitation as well as pain. Isoflurane's ability to suppress behavioural responses does not demonstrate superior analgesia. The study therefore supports lower recorded opioid requirements under the tested pathway, while leaving the adequacy of pain relief and longer-term consequences as separate questions. Other sedatives, including alpha-2 agonists, were restricted, further distinguishing the comparison from some contemporary multimodal practice.

The extubation result also needs its correct time origin. It measured recovery after a decision to terminate sedation for extubation, among a subset selected after randomisation. It did not measure ventilation duration in all randomised patients. A difference of approximately 20 minutes between the medians may be operationally useful, while the longer tail in the midazolam group is also relevant. However, selection for an awakening attempt, unmasked clinical decisions, censoring and re-sedation complicate causal interpretation. Faster emergence after drug cessation should not be presented as proven earlier liberation from intensive care.

Safety conclusions should be more restrained than the statement that no serious event was considered treatment-related. Serious events were common, and several specific event rates were numerically higher with isoflurane. The ten-versus-zero acute kidney injury imbalance is especially important to retain, even though investigators found no evidence linking those events to treatment. In a small trial with many comparisons, chance and baseline clinical differences are plausible explanations. Attribution by investigators cannot, however, establish absence of a causal contribution. The study is insufficient to exclude uncommon or moderately increased risks.

Safety classification itself was imperfect. The main adverse-event table states that the sponsor's categories for events of special interest were not reconciled with the definitions specified for the study, with some events assigned inconsistently. This does not invalidate the listed events, but limits confidence in broad category totals. Participant-level event counts and the explicit descriptions of hypotension, kidney injury and treatment discontinuation are more informative than a general claim of equivalent tolerability.^{1,3}

Applicability is narrowest for the youngest and most complex patients. No child younger than three years was enrolled; many patients in paediatric intensive care fall in that age range. Children with shock, active seizures, suspected raised intracranial pressure or an inability to undergo reliable behavioural scoring were excluded. Exposure was generally brief, with limited information beyond 48 hours. The study cannot establish safety during prolonged sedation, developmental vulnerability in infancy or the efficacy of isoflurane as rescue treatment in the excluded neurological and respiratory conditions.

Correspondence published online in November 2025 questioned the comparator, the predominance of deep sedation and the move from regulatory efficacy to first-line treatment. The investigators defended midazolam as a relevant comparator and emphasised rapid elimination and lower opioid exposure.^{4,5,6} Both positions identify legitimate questions. Non-inferiority to a commonly used drug is useful evidence, but does not establish superiority to alternative sedation pathways. Equally, concern about developmental toxicity should not be presented as clinical harm proved by this trial; long-term neurodevelopment was not adequately assessed.

Sedana Medical funded the study, collaborated in its design and sponsor oversight, and participated in manuscript revision and the decision to publish. Analysis was performed by a contract research organisation, and several investigators had manufacturer relationships, including two employees. These interests do not negate the findings, but support careful separation of measured effects from broader adoption claims. Equipment, training, scavenging, cost and environmental consequences also require local assessment; the trial did not directly establish a complete economic or environmental advantage.

Body of Evidence

The adult Sedaconda trial randomised 301 invasively ventilated patients to isoflurane or propofol for approximately 48 hours. Adequate sedation occupied about 91% of time in both groups, supporting non-inferiority on its chosen endpoint. Isoflurane reduced opioid dose intensity and shortened median awakening time after treatment on day 2. This trial established feasibility and short-term sedative efficacy in adults. Differences in behavioural scales, comparator, patient age and illness make it supportive rather than substitutive evidence for children.²

SLEEPS compared clonidine with midazolam, both accompanied by morphine, in UK paediatric intensive care units. Recruitment fell well below the planned 1,000 children: 129 were randomised and 120 contributed to the primary outcome. Approximately one-third in each group achieved the specified threshold for adequate sedation. Formal

equivalence was not established, although the interval supported the investigators' non-inferiority interpretation. The trial illustrates both the difficulty of recruitment and the limitations of existing sedative regimens; it does not identify a universally optimal alternative to midazolam.⁷

RESTORE studied a broader care strategy in 2,449 children ventilated for acute respiratory failure. A nurse-led, goal-directed sedation pathway did not shorten ventilation compared with usual care: median duration was 6.5 days in both groups. Exploratory findings included less opioid exposure and more time awake and calm, alongside more reported pain and agitation. This large trial demonstrates why achieving less sedation or using fewer drugs must be considered together with comfort and patient-centred outcomes.⁸

The 2025 SESAR trial provides a caution about assuming that inhaled sedation has a uniform clinical benefit. Among 687 adults with moderate-to-severe acute respiratory distress syndrome, sevoflurane rather than propofol produced fewer ventilator-free days and lower 90-day survival. These patients received a different volatile agent, had different disease and generally longer exposure than children in IsoCOMFORT. SESAR therefore does not establish that isoflurane is harmful in this paediatric setting. It does show that pharmacological plausibility and satisfactory sedation are insufficient substitutes for trials measuring recovery and survival.⁹

Long-term developmental reassurance requires similarly careful boundaries. The GAS trial randomised 722 infants undergoing hernia repair to awake regional or sevoflurane-based general anaesthesia. With a median general-anaesthetic duration of 54 minutes, the five-year intelligence outcome supported equivalence within the prespecified margin. This is reassuring for that brief exposure in a selected surgical population. It cannot resolve the developmental effects of days of sedation in critically ill children, or establish the safety of prolonged isoflurane exposure.¹⁰

IsoCOMFORT adds a useful but limited piece to this evidence. It supports the feasibility of isoflurane as an alternative for selected children aged three years or older, with lower recorded opioid exposure and rapid emergence. Its borderline primary non-inferiority result, failed per-protocol confirmation and small safety population do not establish that routine first-line substitution improves outcomes. The next clinical comparison must retain reliable comfort assessment while measuring ventilation, complications, recovery and longer-term development.

Should inhaled isoflurane replace intravenous midazolam as routine first-line sedation in ventilated children?

IsoCOMFORT supports a possible alternative in selected children aged 3–17 years, but its non-inferiority finding was fragile. It does not establish routine first-line superiority or safety during prolonged exposure.

References

1. Miatello J, Palacios-Cuesta A, Radell P, Oberthuer A, Playfor S, Amores-Hernández I, et al; IsoCOMFORT Study Group. Inhaled isoflurane for sedation of mechanically ventilated children in intensive care (IsoCOMFORT): a multicentre, randomised, active-control, assessor-masked, non-inferiority phase 3 trial. *Lancet Respir Med*. 2025;13(10):897-910.
2. Meiser A, Volk T, Wallenborn J, Guenther U, Becher T, Bracht H, et al. Inhaled isoflurane via the anaesthetic conserving device versus propofol for sedation of invasively ventilated patients in intensive care units in Germany and Slovenia: an open-label, phase 3, randomised controlled, non-inferiority trial. *Lancet Respir Med*. 2021;9(11):1231-1240.
3. Miatello J, et al; IsoCOMFORT Study Group. Supplementary appendix to the IsoCOMFORT trial. *Lancet Respir Med*. 2025.
4. Engel J, Fideler F, Nordmeyer J, Drexler B, Neunhoeffler F. Rethinking isoflurane for paediatric ICU sedation. *Lancet Respir Med*. Published online November 10, 2025.
5. Bragg EA, Nadkarni AS, Barnes SS, Fackler JC, Kudchadkar SR. Rethinking isoflurane for paediatric ICU sedation. *Lancet Respir Med*. Published online November 10, 2025.
6. Miatello J, Palacios-Cuesta A, Radell P, Trieschmann U, Sackey P, Tissieres P. Rethinking isoflurane for paediatric ICU sedation - Authors' reply. *Lancet Respir Med*. Published online November 10, 2025.
7. Wolf A, McKay A, Spowart C, Granville H, Boland A, Petrou S, et al. Prospective multicentre randomised, double-blind, equivalence study comparing clonidine and midazolam as intravenous sedative agents in critically ill children: the SLEEPS (Safety profile, Efficacy and Equivalence in Paediatric intensive care Sedation) study. *Health Technol Assess*. 2014;18(71):1-212.

8. Curley MA, Wypij D, Watson RS, Grant MJ, Asaro LA, Cheifetz IM, et al. Protocolized sedation vs usual care in pediatric patients mechanically ventilated for acute respiratory failure: a randomized clinical trial. *JAMA*. 2015;313(4):379-389.
9. Jabaudon M, Quenot JP, Badie J, Audard J, Jaber S, Rieu B, et al. Inhaled Sedation in Acute Respiratory Distress Syndrome: The SESAR Randomized Clinical Trial. *JAMA*. 2025;333(18):1608-1617.
10. McCann ME, de Graaff JC, Dorris L, Disma N, Withington D, Bell G, et al. Neurodevelopmental outcome at 5 years of age after general anaesthesia or awake-regional anaesthesia in infancy (GAS): an international, multicentre, randomised, controlled equivalence trial. *Lancet*. 2019;393(10172):664-677.

Kids THRIVE

George S, Williams T, Humphreys S, Atkins T, Tingay D, Gelbart B, et al. Effectiveness of nasal high-flow oxygen during apnoea on hypoxaemia and intubation success in paediatric emergency and ICU settings: a randomised, controlled, open-label trial. *Lancet Respir Med.* 2025;13(6):545–555.

Introduction

Emergency intubation in children combines limited physiological reserve with a technically demanding procedure. Oxygen desaturation can develop rapidly while ventilation is interrupted, particularly in small infants or children with respiratory disease. Avoiding hypoxaemia and achieving successful intubation at the first attempt are related objectives, but they are not interchangeable. Oxygen delivery may preserve saturation without improving the laryngeal view, tube passage or the operator's ability to complete the procedure.¹

Nasal high-flow oxygen offers a way to continue supplying oxygen during laryngoscopy after a preoxygenation mask has been removed. The technique is attractive because nasal cannulae can remain in place while the mouth is accessible. Earlier physiological work in healthy anaesthetised children demonstrated a longer period before desaturation, providing a rationale for emergency use.² Its effectiveness in an acutely ill child may nevertheless depend on airway patency, lung disease, prior oxygenation, haemodynamics and whether equipment can be applied promptly.

Three distinct decisions can easily become conflated: how to preoxygenate before induction, whether to continue oxygen during apnoea, and how to achieve first-attempt intubation. A trial changing only the second component cannot establish the best overall airway strategy. Equally, a favourable physiological result in an elective setting does not establish benefit during emergency intubation across neonatal, paediatric intensive care and emergency departments.

Kids THRIVE addressed this uncertainty by randomising emergency intubation encounters to nasal high-flow oxygen during apnoea or standard care without supplemental oxygen during that period. Its neutral primary comparison, favourable per-protocol hypoxaemia result and neonatal subgroup require careful separation. They answer different questions and do not have equal protection from bias.¹

Synopsis

Kids THRIVE was an investigator-led, randomised, open-label trial conducted at ten hospitals in Australia, New Zealand and Switzerland. Participating locations comprised ten paediatric intensive care units, four emergency departments and one non-maternity neonatal intensive care unit. Recruitment took place between May 2017 and October 2022. Children younger than 16 years requiring emergency endotracheal intubation were eligible. Exclusions included planned nasal intubation, endotracheal tube exchange, immediate intubation during cardiorespiratory arrest and anatomical or traumatic obstruction preventing nasal oxygen delivery.^{1,3}

The unit of allocation was an intubation encounter. A child could therefore participate more than once. Sequentially numbered, opaque, sealed envelopes provided 1:1 allocation, stratified by age, operator experience and location. Consent could be obtained prospectively or through an approved consent-to-continue process. Most encounters were randomised before consent was sought. Clinicians and outcome assessors knew treatment assignment.

Preoxygenation, induction drugs, neuromuscular blockade and other airway management remained at the treating team's discretion. In the intervention group, heated humidified nasal oxygen was applied when preoxygenation ended and the child became apnoeic. Weight-based flow settings used approximately 2 L/kg/min in smaller children and fixed higher flows in larger children, up to 50 L/min. The intended inspired oxygen fraction was generally 1.0, with adjustment for cyanotic heart disease. The comparator received no supplemental oxygen during the apnoeic phase. This was not a comparison of high-flow with low-flow nasal oxygen, and it did not randomise the preoxygenation method.

The two primary outcomes concerned the first intubation attempt: hypoxaemia and successful intubation without desaturation. Hypoxaemia was defined as peripheral oxygen saturation of 90% or less, or a fall of at least ten percentage points in children with a baseline saturation below 90% or cyanotic intracardiac right-to-left shunting. An attempt extended from insertion to removal of the laryngoscope. Successful tube placement required confirmation using exhaled carbon dioxide. Video recordings of the procedure and physiological monitor were used to verify the clinical record where available.

The investigators planned approximately 960 encounters, allowing for exclusions and missing data. The hypoxaemia calculation assumed a reduction from 16% to 8%, a substantial relative effect. Primary analyses used mixed-effects logistic regression, incorporating patient and hospital effects and adjustment for age and operator

experience. The two co-primary outcomes were assessed using 97.5% confidence intervals to account for multiplicity. A modified intention-to-treat analysis included consented encounters with available primary outcomes; a per-protocol analysis excluded encounters that did not receive the allocated intervention.

Of 1,069 randomised encounters in 969 children, 950 encounters in 860 children entered the primary analysis: 476 assigned high-flow and 474 standard care. The difference reflects withdrawals, consent and eligibility exclusions, missing primary outcomes and one encounter in which intubation did not proceed. Baseline and several secondary-outcome tables used a larger consented population of 480 versus 481 encounters. These denominators should not be treated as interchangeable with either the primary population or the number of individual children.

The median age was approximately six months, and around 59% of encounters involved children younger than one year. About four fifths occurred in paediatric intensive care. Junior clinicians performed approximately three quarters of first attempts, and neuromuscular blockade was administered in about 95%. Median saturation immediately before the procedure was 100% in both groups. High-flow support before intubation was more common in the intervention group, whereas video laryngoscopy was more common in controls. Video recordings were obtained for approximately 95% of consented encounters.

Hypoxaemia occurred in 61 of 476 high-flow encounters, 12.8%, and 77 of 474 controls, 16.2%; the adjusted odds ratio was 0.74 (97.5% CI 0.46–1.18; $P=0.15$). Successful first-attempt intubation without desaturation occurred in 300 encounters, 63.0%, versus 280, 59.1%; the adjusted odds ratio was 1.13 (97.5% CI 0.79–1.62; $P=0.43$). Neither primary comparison demonstrated superiority.¹

The per-protocol population excluded 32 high-flow and 13 control encounters from the primary population. Hypoxaemia then occurred in 48 of 444 encounters, 10.8%, versus 77 of 461, 16.7%; the adjusted odds ratio was 0.59 (97.5% CI 0.36–0.97; $P=0.017$). First-attempt success without desaturation remained inconclusive: 64.0% versus 58.1%, adjusted odds ratio 1.22 (97.5% CI 0.87–1.71). In the neonatal intensive care subgroup, hypoxaemia occurred in 30.0% versus 57.1%, but the interaction test did not establish that treatment effects differed by location.

Secondary outcomes showed no clear improvement in the number of intubation attempts, need for reoxygenation, duration of ventilation or length of stay. Median ventilation duration was approximately 71 hours in both groups. A sensitivity analysis using saturation below 80%, excluding children with intracardiac shunting, also remained

inconclusive. Minor and major intubation-related adverse events were distributed across both groups. Cardiac arrest occurred in four high-flow and two control encounters; no deaths occurred within 60 minutes. Hospital deaths numbered 36 and 48, without a demonstrated mortality effect. The trial was publicly and charitably funded; Fisher & Paykel supplied equipment and consumables without charge.

Critique

The primary interpretation is that adding high-flow oxygen during apnoea did not demonstrate improved emergency intubation outcomes in this population. That conclusion should neither be strengthened into proof of no physiological effect nor displaced by the positive per-protocol result. The hypoxaemia confidence interval includes an appreciable benefit as well as no benefit. The trial had been designed around halving hypoxaemia, so smaller useful effects remain uncertain. There was also no convincing improvement in the more integrated endpoint of first-attempt success without desaturation.

The distinction between randomised allocation and actual treatment is central. Removing non-adherent encounters eliminated 13 hypoxaemic events from the high-flow group and none from controls. Calculated from the published counts, 13 of the 32 excluded high-flow encounters, 40.6%, had experienced hypoxaemia. Their removal therefore changed both the population and its event rate. This does not show that exclusion was chosen to produce a favourable result; it shows why conditioning on treatment delivery can undermine the original comparability of the groups.^{1,3}

Non-adherence was not uniformly an administrative accident. The supplementary table records patient deterioration in eight high-flow encounters, alongside omissions, clinician decisions, changes in the intended intubation route and equipment difficulties. Clinical circumstances that make high-flow difficult to apply may also increase the risk of desaturation. A per-protocol comparison cannot distinguish the effect of receiving oxygen from the effect of being sufficiently stable and straightforward to receive it. Consequently, the result does not establish that high-flow would prevent hypoxaemia if every clinician simply complied.

The accompanying editorial appropriately highlights implementation barriers and the divergence between analyses.⁴ However, suggesting that benefit is established wherever clinicians are willing and able to deliver high-flow goes beyond the randomised evidence. Better implementation might increase benefit, but that remains a plausible explanation to test. A future implementation study would need to improve reliable

delivery while preserving an unbiased comparison, including the difficult encounters most likely to be excluded by an adherence-based analysis.

Even the modified intention-to-treat population did not retain every randomised encounter. Overall, 119 of 1,069 encounters were absent from the primary analysis. Some exclusions were unavoidable consequences of emergency consent or missing recordings, but clinician withdrawal before intubation was more common after high-flow allocation: 19 versus five encounters. If those decisions were related to instability or anticipated difficulty, selection could influence the estimate. The eleven encounters with unavailable primary outcomes alone were a small proportion; focusing only on those eleven would understate the larger post-randomisation exclusion problem.

The study nevertheless achieved unusually detailed procedural ascertainment for this setting. Video capture allowed assessment of timing, saturation and intervention delivery rather than reliance solely on retrospective recollection. Repeated encounters in the same child and differences between hospitals were addressed in the statistical model. Adjustment for the two primary comparisons reduced the likelihood that a chance positive endpoint would determine the overall conclusion. These features strengthen the reported primary comparison without removing the limitations introduced by exclusion and open treatment.

Masking was impractical, but the consequences extend beyond measurement of an objective saturation value. Knowing that oxygen was being delivered could influence how long an operator persisted, when help was requested or when an attempt was abandoned. The first-attempt composite depends on these decisions as well as oxygenation. Both groups had substantial numbers of attempts abandoned without desaturation. Such abandonment is not necessarily inappropriate: inability to see the larynx, difficulty passing the tube or other hazards may justify stopping before oxygen saturation falls. The trial cannot support prolonging an attempt merely because high-flow is present.

Preoxygenation is another important boundary. Most encounters used bag-mask ventilation beforehand, but technique and adequacy were not standardised, and expired oxygen concentrations were not measured as a routine study endpoint. A saturation of 100% does not establish that the lungs are fully denitrogenated or that oxygen reserve is equivalent. The supplied correspondence raises this issue, and the investigators acknowledge that some patients probably received suboptimal preoxygenation.^{5,6} This limits the physiological explanation of the result; it does not turn the trial into a comparison of preoxygenation strategies.

The neonatal result is clinically interesting but statistically insufficient to establish a distinct responsive subgroup. The high-flow estimate within the neonatal unit was favourable, yet the interaction P value for location was 0.11. A significant result in one subgroup and a non-significant result in another does not itself demonstrate different treatment effects. Moreover, the neonatal unit was a single non-maternity site with a particular case mix. Its patients differed from neonates managed in paediatric intensive care, including their frequency of prematurity. Location, developmental maturity and clinical indication cannot be separated by this small comparison.

Outcome definitions also matter when interpreting apparent agreement with other studies. The primary hypoxaemia endpoint used a 90% threshold or a relative change appropriate to children with low baseline saturation. It is different from severe desaturation below 80%, and first-attempt success without desaturation is different from technical first-pass success alone. The severe-desaturation analysis excluded children with shunting and involved a smaller population. Presenting its rate as if it applied to all randomised children would overstate comparability with other airway cohorts.

Several features constrain transportability. The cohort was dominated by infants, paediatric intensive care, neuromuscular blockade and junior first operators. Cardiorespiratory arrest and nasal intubation were excluded. Approximately 730 eligible patients were not approached, and the study does not establish whether enrolment was equally feasible during the most pressured emergencies. Video laryngoscopy was used in only about one third of encounters and was imbalanced between groups. Its effect within this trial was not randomised, so comparisons by laryngoscope should not be used to infer which device caused better or worse outcomes.

The adverse-event findings provide useful observation but limited reassurance about uncommon serious harm. Six arrests and no immediate deaths cannot establish safety equivalence. Event totals are not counts of unique affected children, and some participants contributed more than one encounter. Similarly, the absence of an improvement in ventilation duration or hospital stay does not rule out a procedural oxygenation benefit, but it prevents such a benefit from being assumed to translate into faster recovery.

Overall, Kids THRIVE challenges routine adoption based on a presumed universal emergency-intubation benefit. It also leaves room for a modest effect and for population-specific benefits supported by other evidence. Equipment, preparation and staff familiarity remain relevant, but high-flow should be judged as one component of a complete airway strategy. The most defensible conclusion follows the primary allocation

comparison, with the adherence and neonatal signals used to guide further investigation.

Body of Evidence

The development of apnoeic oxygenation illustrates why favourable physiology requires testing in the intended clinical setting. In the 2017 THRIVE trial, 48 healthy children undergoing elective anaesthesia were randomised to age-specific nasal high-flow or control after standardised preoxygenation. High-flow prolonged the time during which oxygen saturation was maintained. Carbon dioxide increased at a similar rate in both groups, showing that preserved oxygenation did not establish effective ventilation.² These controlled conditions differ substantially from an unstable child with diseased lungs undergoing urgent intubation.

Evidence from critically ill adults had already demonstrated that extending elective findings was uncertain. The 150-patient FELLOW trial compared nasal oxygen at 15 L/min during laryngoscopy with no supplemental oxygen during that period. Median lowest saturation was 92% versus 90%, without a clear difference, and severe desaturation and clinical outcomes were also inconclusive.⁷ This was a different population and delivery system from Kids THRIVE, but it reinforces the distinction between physiological potential and demonstrated effectiveness in critical illness.

The most directly relevant positive neonatal evidence came from the 2022 SHINE trial. At two Australian neonatal intensive care units, 251 intubations in 202 infants were analysed after allocation to nasal high-flow or standard care. Median postmenstrual age was 27.9 weeks and median weight 920 g. First-attempt success without physiological instability occurred in 50.0% versus 31.5%, with an adjusted absolute difference of 17.6 percentage points (95% CI 6.0–29.2). First-attempt success regardless of physiological stability was also higher with high-flow.⁸

SHINE supplies randomised support in a small, predominantly premature neonatal population. Its composite incorporated a substantial fall from baseline saturation or bradycardia, rather than the Kids THRIVE definition. The neonatal signal in Kids THRIVE is consistent with the possibility of benefit in such patients, but is not an independent, adequately powered replication of SHINE across neonatal practice. Conversely, the overall Kids THRIVE result does not invalidate the more focused neonatal trial.

Other airway interventions address different causes of failure. In the 2024 VODE trial, 214 neonates with analysed outcomes had been randomised to video or direct laryngoscopy. First-attempt success occurred in 74% versus 45%, respectively.⁹ This

directly tested visualisation rather than oxygen delivery and used technical first-attempt success as its primary endpoint. It helps explain why supplying oxygen alone may have limited influence when the principal obstacle is seeing the airway or placing the tube.

The 2024 PREOXI trial separately tested preoxygenation in 1,301 critically ill adults. Non-invasive ventilation reduced hypoxaemia below 85% during intubation compared with an oxygen mask: 9.1% versus 18.5%, an absolute difference of –9.4 percentage points (95% CI –13.2 to –5.6).¹⁰ This provides strong evidence for a different intervention at a different stage of the procedure. It cannot establish the best paediatric preoxygenation method or explain away Kids THRIVE, but it emphasises why preparation of oxygen reserve and oxygen delivery during apnoea should be evaluated separately.

By the end of 2025, the evidence therefore supported a context-dependent assessment. Nasal high-flow can prolong maintained oxygenation during elective anaesthesia and has demonstrated neonatal benefits in a focused trial. Kids THRIVE did not establish improved outcomes from routine addition during emergency intubation across its broader paediatric population. The remaining questions concern reliable delivery, the size of any modest average benefit, and which clinical settings combine sufficient physiological responsiveness with practical feasibility.

Should nasal high-flow oxygen be added routinely during apnoea for emergency intubation in children?

Kids THRIVE does not establish a routine benefit across this population. Its per-protocol result should not override the primary analysis; neonatal use should also consider the separate neonatal trial evidence.

References

1. George S, Williams T, Humphreys S, Atkins T, Tingay D, Gelbart B, et al. Effectiveness of nasal high-flow oxygen during apnoea on hypoxaemia and intubation success in paediatric emergency and ICU settings: a randomised, controlled, open-label trial. *Lancet Respir Med*. 2025;13(6):545-555.
2. Humphreys S, Lee-Archer P, Reyne G, Long D, Williams T, Schibler A. Transnasal humidified rapid-insufflation ventilatory exchange (THRIVE) in children: a randomized controlled trial. *Br J Anaesth*. 2017;118(2):232-238.

3. George S, Williams T, Humphreys S, et al. Supplementary appendix to: Effectiveness of nasal high-flow oxygen during apnoea on hypoxaemia and intubation success in paediatric emergency and ICU settings. *Lancet Respir Med*. 2025;13:545–555.
4. Alexander EC, Ramnarayan P. Nasal high-flow oxygen for apnoeic oxygenation during emergency endotracheal intubation of children. *Lancet Respir Med*. 2025;13(6):480-482.
5. Pal N, Tran BW. What is adequate preoxygenation?. *Lancet Respir Med*. 2025;13(7):e35.
6. George S, Williams T. What is adequate preoxygenation? - Authors' reply. *Lancet Respir Med*. 2025;13(7):e36.
7. Semler MW, Janz DR, Lentz RJ, Matthews DT, Norman BC, Assad TR, et al. Randomized Trial of Apneic Oxygenation during Endotracheal Intubation of the Critically Ill. *Am J Respir Crit Care Med*. 2016;193(3):273-280.
8. Hodgson KA, Owen LS, Kamlin COF, Roberts CT, Newman SE, Francis KL, et al. Nasal High-Flow Therapy during Neonatal Endotracheal Intubation. *N Engl J Med*. 2022;386(17):1627-1637.
9. Geraghty LE, Dunne EA, Ní Chathasaigh CM, Vellinga A, Adams NC, O'Curraín EM, et al. Video versus Direct Laryngoscopy for Urgent Intubation of Newborn Infants. *N Engl J Med*. 2024;390(20):1885-1894.
10. Gibbs KW, Semler MW, Driver BE, Seitz KP, Stempek SB, Taylor C, et al. Noninvasive Ventilation for Preoxygenation during Emergency Intubation. *N Engl J Med*. 2024;390(23):2165-2177.

REDvent

Khemani RG, Bhalla A, Hotz JC, Klein MJ, Kwok J, Kohler K, et al. Randomized Trial of Lung and Diaphragm Protective Ventilation in Children. NEJM Evid. 2025;4(6):EVIDoa2400360.

Introduction

Mechanical ventilation can protect the injured lung while leaving the respiratory muscles almost inactive. Reducing support encourages diaphragmatic activity, but excessive effort can itself increase lung stress and cause muscle injury. The challenge is therefore to provide sufficient assistance for gas exchange and lung protection without suppressing respiratory activity unnecessarily. This balance changes as illness resolves and is difficult to maintain through occasional bedside assessments alone.

The attraction of effort-guided ventilation is that assistance could be adjusted to measured respiratory work rather than inferred from airway pressure and tidal volume. A small adult randomised study showed that titrating support to a specified range of diaphragmatic effort substantially increased the proportion of breaths within that range. It established physiological feasibility, but did not establish an improvement in survival, recovery or duration of ventilation.¹

Children create additional challenges because respiratory rate, airway resistance, developmental physiology and the reasons for prolonged ventilation vary widely. A phase I study of Real-Time Effort-Driven ventilator management, REDvent, combined computer decision support with oesophageal manometry in 32 children. The intervention was feasible and outcomes compared favourably with matched historical controls. Historical comparisons could not reliably separate treatment effects from changes in practice or patient selection, making a randomised evaluation necessary.²

The phase II REDvent trial tested a combined lung- and diaphragm-protective strategy in children with acute respiratory distress syndrome. It incorporated two randomisations, one early during ventilation and another among patients who failed their first spontaneous breathing trial. This allowed investigation of both the acute management package and a later effort-guided weaning intervention, although the two comparisons answer different questions.³

Synopsis

REDvent was an open-label, single-centre phase II randomised trial at Children's Hospital Los Angeles. Recruitment took place from late 2017 to March 2024. Eligible patients were older than one month and no older than 21 years, had received invasive ventilation for less than 72 hours, and met hypoxaemic and radiographic criteria for paediatric acute respiratory distress syndrome. An oxygenation index of at least 4 or oxygen saturation index of at least 5, with pulmonary infiltrates, was required. Contraindications to oesophageal manometry and conditions preventing conventional weaning, including intracranial hypertension, a critical airway or extracorporeal support, were exclusions.³

The acute-phase randomisation allocated patients equally to REDvent or usual care, stratified by age category and immunosuppression. Oesophageal balloon catheters were placed in both groups to permit physiological measurements. In REDvent, the measurements also informed bedside treatment. Research physiological measurements in the usual-care group were not shared with the clinical team. The intervention was delivered through a computer decision-support system that recommended changes; it did not autonomously control the ventilator.^{3,4}

REDvent used synchronised intermittent mandatory ventilation with pressure control and pressure support. Respiratory therapists interacted with the system every four hours. Oxygenation recommendations followed the ARDS Network low-PEEP/inspired-oxygen table, targeting oxygen saturation of 88–93% or arterial oxygen tension of 55–68 mmHg. Ventilator rate and inspiratory pressure were reduced when pH allowed, while pressure limits and permissive hypercapnia constrained escalation. These components provided a structured lung-protective framework in addition to the effort-guided component.

Oesophageal manometry estimated the pressure swing generated by the patient's inspiratory effort. The pressure-rate product was calculated as the magnitude of that swing multiplied by respiratory rate. REDvent sought a range of 200–400 breaths/min × cm H₂O. Low effort could prompt less assistance, while excessive effort could prompt more assistance within pressure limits. Respiratory therapists could implement accepted recommendations immediately; disagreement prompted discussion with the attending physician and documentation of the decision. The package combined measurement, software, regular reassessment and authority to act.⁴

Usual care followed general lung-protective principles without an explicit computer or paper algorithm. Respiratory therapists still performed routine four-hourly ventilator checks, but changes other than inspired oxygen generally required a physician decision. Sedation, fluid treatment and other cointerventions remained with the clinical team.

Both groups underwent the same daily screening for readiness to attempt a spontaneous breathing trial, reducing differences attributable simply to whether readiness was sought.

Screening included spontaneous breathing, appropriate pH, PEEP no higher than 10 cm H₂O and inspired oxygen no higher than 0.6. Patients then underwent a 30-minute oxygenation test with PEEP of 5 cm H₂O and inspired oxygen no higher than 0.5. Those maintaining oxygen saturation above 90% proceeded to a two-hour spontaneous breathing trial on continuous positive airway pressure of 5 cm H₂O without pressure support. Failure criteria were standardised and included excessive pressure-rate product. Passing the trial satisfied the primary endpoint even if extubation was delayed for other reasons.^{3,4}

Patients who failed their first breathing trial and were not extubated within six hours entered a second randomisation, stratified by age and acute-phase allocation. REDvent-weaning used pressure support or continuous positive airway pressure, with support adjusted every four hours towards the same effort target. The comparator was clinician-directed weaning. Daily breathing trials continued in both groups. This second comparison concerned a selected population that had already failed an initial liberation assessment.

The primary outcome was length of weaning: time from initiation of the first spontaneous breathing trial until trial passage or successful extubation, whichever occurred first. It was not total duration of mechanical ventilation. Secondary outcomes included time to breathing-trial passage, ventilation duration among survivors, ventilator-free days, mortality, inspiratory strength, reintubation, non-invasive support and functional status.³

The analysis changed during the study. Initial planning used a time-to-event model and a target sample of 300. Slower recruitment during the pandemic prompted review of the pooled outcome distribution without reference to treatment allocation. The analysis was revised to negative binomial regression and the minimum sample to 230. During final analysis, two unusually prolonged weaning courses strongly influenced that model. The published primary comparison therefore used an ordinal model with six categories, from less than half a day to at least 7.5 days. Adjustment included age, immunosuppression and baseline oxygenation index. Secondary confidence intervals were not corrected for multiple comparisons.³

Of 340 eligible patients, 248 were randomised: 126 to REDvent and 122 to usual care. One REDvent patient withdrew before trial procedures, leaving 125 and 122 in the main

dataset. The primary outcome was analysed in 109 and 106 patients. The flow diagram records early deaths, evolving contraindications, limitations of treatment and other reasons for stopping acute-phase procedures. All retained patients contributed to the broader secondary-outcome dataset, although individual measures had different denominators.^{3,4}

Median age was 5.9 versus 4.9 years. Approximately four-fifths had a comorbidity and just over a quarter were immunocompromised. Pneumonia was the most common trigger. Randomisation occurred at about 20 hours after intubation. The REDvent group contained somewhat more patients with severe oxygenation impairment and more receiving inhaled nitric oxide at baseline; this contributed to the decision to adjust for oxygenation index. The population represented a medically complex tertiary paediatric intensive care cohort.

Implementation was strong: 91.4% of oxygenation and 94.9% of ventilation recommendations were followed during the acute phase. Both groups had high adherence to eligible breathing trials. When patients were breathing spontaneously, REDvent reduced peak inspiratory pressure by an adjusted mean of 3 cm H₂O, plateau pressure by 1.9 cm H₂O and PEEP by 2 cm H₂O. Median oesophageal pressure swing was -8.3 versus -6.7 cm H₂O. The more negative swing indicates greater patient effort, not less effort. Tidal volume and airway and transpulmonary driving pressures were similar.^{3,4}

For the primary outcome, 60/109 REDvent patients and 41/106 controls had a weaning duration below half a day: 55.0% versus 38.7%. The adjusted odds ratio for being in a shorter weaning category was 1.67, 95% confidence interval 1.01–2.77; P=0.045. The negative binomial analysis also favoured REDvent, with an incidence rate ratio of 0.54, but was strongly influenced by two controls who had not passed a breathing trial by 90 days.³

Median time from intubation to successful breathing-trial passage was 3.83 versus 4.75 days, with an adjusted relative estimate of 0.80. Ventilation duration among survivors was 5.0 versus 5.6 days, adjusted relative estimate 0.81, 95% confidence interval 0.66–0.98. These ratios are modelled relative effects and should not be confused with absolute differences in days. The proportion passing the first breathing trial was 59/110 versus 39/107, 53.6% versus 36.5%, distinct from the primary endpoint's shortest category.³

Median ventilator-free days were 21.9 versus 21.3, without a clear difference. Ninety-day mortality was 21/125 versus 18/122, adjusted odds ratio 1.04, 95% confidence interval 0.44–2.47. Reintubation counts were eight in each acute-phase group, and non-invasive

respiratory support after extubation was common. Functional status at intensive care discharge favoured REDvent, but this difference was not sustained clearly at hospital discharge. Worsening cerebral performance category at hospital discharge occurred in nine versus three patients. These were exploratory secondary observations.³

The weaning-phase randomisation allocated 58 patients to each group. It did not show a clear improvement in the ordinal weaning outcome: adjusted odds ratio 1.3, 95% confidence interval 0.7–2.6. Adverse events did not establish a clear difference overall. Oesophageal catheter misplacement into the tracheobronchial tree was recorded in three REDvent patients, illustrating that the monitoring procedure itself requires expertise. The study was too small to establish safety for uncommon complications.⁴

Critique

REDvent offers an encouraging efficacy signal and a convincing demonstration that a complex ventilation package can be delivered with high adherence. Its main contribution is the randomised comparison after an earlier historical-control pilot. The intervention changed actual bedside ventilation and respiratory effort, and the pattern of earlier breathing-trial success and shorter ventilation among survivors is clinically coherent. These findings justify a larger trial, but they do not yet establish a broadly applicable treatment standard.

The primary outcome deliberately targets the transition from the first breathing trial to physiological readiness for extubation. That can be useful in a phase II study because it avoids delays caused by procedures, feeds or airway concerns after respiratory readiness has been achieved. It also creates distance from the outcome patients experience. A child can meet the primary endpoint while remaining intubated, and children who die or never become eligible for a trial are absent from that comparison. Faster completion of this selected phase is not necessarily equivalent to more survival free of ventilation.

Only 215 of the 247 retained patients entered the primary analysis. Selection occurred after randomisation and was related to the clinical course, including death and evolving contraindications. Randomisation balances groups at allocation; it does not guarantee balance within the subset who later reach a particular milestone. The broadly similar mortality and ventilator-free-day results provide necessary context, although their wide intervals do not prove equivalence. The primary effect is best understood as an effect on observed weaning among assessable patients, rather than a complete account of outcomes for all randomised children.^{3,4}

The statistical revisions also matter. A change prompted by a blinded review of the pooled distribution can be reasonable, particularly when the original model's assumptions are unsuitable. The further change to ordinal analysis after two influential observations became apparent introduces additional analytical flexibility. Both the negative binomial and ordinal models favoured REDvent, which is reassuring about direction, but the headline ordinal confidence interval barely excludes no difference. The evidence is therefore less definitive than a prespecified, unchanged primary analysis replicated in a large trial would be.

An odds ratio of 1.67 does not mean that weaning was 67% faster, nor that ventilation was reduced by a fixed number of days. It describes the odds of occupying a shorter category under the model's proportional-odds assumption. Categorisation reduces the influence of exceptionally long courses but also discards information about differences within each category. Presenting the underlying distribution alongside the model is valuable: much of the apparent advantage was concentrated in patients who completed weaning on the day of their first trial.

The two randomisations should not be collapsed into a simple comparison of early versus late treatment. The second randomisation only included patients who failed their first breathing trial, a group partly shaped by acute-phase treatment. Its 116 participants provided limited precision, and the effort-guided weaning intervention was narrower than the original package. A neutral late-phase result is compatible with benefit from early prevention, but it does not prove that the timing of diaphragm protection explains the difference. Patients, interventions and statistical power all differed.

The mechanism remains unresolved. REDvent simultaneously altered PEEP, inspiratory assistance, frequency of ventilator changes and the workflow for making those changes. Lower ventilator assistance was accompanied by more negative oesophageal pressure swings, consistent with patients taking on more respiratory work. Yet pressure-rate product, tidal volume and driving pressures did not clearly separate to the same extent. Inspiratory strength was greater at the first trial but similar at extubation. These findings support physiological engagement without demonstrating prevention of diaphragmatic atrophy as the causal pathway.³

The intervention could also work through more prompt de-escalation. Usual-care PEEP was often higher than the low-PEEP/oxygen table, whereas REDvent followed it closely. Earlier reductions may make patients eligible for the oxygenation test sooner and reduce avoidable support. This is a valid treatment effect if confirmed, but it need not require every element of the oesophageal-effort algorithm. The accompanying editorial appropriately asked whether regular, empowered clinician or therapist adjustment could

produce similar results. A comparison with equally structured active care would help isolate the additional value of effort measurement and software.⁵

Importantly, usual-care patients did receive routine four-hourly respiratory-therapist checks. The distinction was therefore not regular attention versus neglect. In the intervention group, measurements generated explicit recommendations that therapists could implement, whereas usual-care adjustments generally depended on physician decisions. The workflow, decision frequency and permission to act are part of the intervention. Implementing the same software in a unit with different staffing and professional responsibilities may not reproduce the trial's effect.⁴

The open-label design leaves room for changes in other management and in decisions surrounding extubation. Standardised daily screening, oxygenation tests and breathing-trial criteria reduce this concern, and treatment adherence was high. Nevertheless, the clinical team retained control of extubation, and the supplement records many occasions on which a passed trial was not followed by extubation within six hours. Sedation, procedures and neurological or airway concerns were common reasons. The primary endpoint and actual ventilation duration therefore measure related but distinct processes.

Secondary findings require restraint. Earlier successful trials, shorter survivor ventilation and better functional status at intensive care discharge all favour the intervention, but numerous outcomes were assessed without multiplicity adjustment. The functional signal was weaker by hospital discharge, while cerebral performance deterioration numerically favoured usual care. Renal replacement therapy was less frequent with REDvent, but baseline renal differences and a small number of events complicate causal interpretation. None of these observations should independently establish neurological benefit, renal protection or harm.^{3,4}

External validity is limited by a single specialist centre, its experience with oesophageal manometry and the custom software and support infrastructure. Accurate balloon positioning, calibration and waveform interpretation are not trivial. The broad age range is useful, but the small sample cannot establish effectiveness in every developmental or disease subgroup. Recruitment over more than six years, including the pandemic, further reflects a setting in which clinical practice and admission patterns could vary. Multicentre confirmation should measure patient-centred outcomes and the resources required to deliver the package.

Body of Evidence

The phase I REDvent study enrolled 32 children and compared them with 128 matched historical controls. Acceptance of recommendations exceeded 75%, and the intervention was associated with six more ventilator-free days and three fewer days of ventilation among survivors. These large apparent effects were hypothesis-generating because the comparison was non-randomised. The phase II result is a more credible estimate of the package's potential, with a smaller difference in ventilation duration and no clear increase in ventilator-free days.²

In the adult trial by de Vries and colleagues, 40 patients receiving partial ventilatory support were randomised to hourly titration based on diaphragmatic effort or standard care. Over 24 hours, 81% versus 35% of breaths fell within the specified effort range, without clear differences in tidal volume or dynamic transpulmonary pressure. REDvent extends this physiological line of investigation towards clinical outcomes in children. Different effort measurements, targets and populations mean that the adult study cannot independently validate REDvent's clinical effect.¹

The SANDWICH trial evaluated a broader sedation and ventilator-liberation protocol across 18 paediatric intensive care units, including 8,843 children. Time to successful extubation was modestly reduced, with an adjusted hazard ratio of 1.11. Its much larger scale illustrates that changes to systematic readiness assessment and clinical workflow can affect liberation, even without an oesophageal-effort target. REDvent's usual-care group already received standardised trials, so the two studies tested different additions to existing practice.⁶

RESTORE randomised paediatric intensive care units to a nurse-led sedation protocol or usual care and included 2,449 children with acute respiratory failure. Median ventilation duration was 6.5 days in both groups. Although sedative exposure and wakefulness differed, these changes did not translate into shorter ventilation. Together, RESTORE and SANDWICH caution against assuming that every sensible protocol improves liberation; the content of the intervention, comparator and local practice determines its effect.⁷

A 21-unit before-and-after study of lung-protective ventilation in paediatric acute respiratory distress syndrome included 693 patients. Protocol implementation improved adherence but did not clearly reduce 60-day mortality. Associations between greater adherence and better outcomes were observational and could reflect differences in patients as well as care. REDvent's randomisation is a methodological advance, while its small size and combined intervention leave open which components are necessary for a larger clinical benefit.⁸

Work within the REDvent programme has also explored airway occlusion measurements as less invasive surrogates for effort. An analysis of 97 children found expiratory occlusion pressure more closely correlated with oesophageal pressure swing than other bedside measures. This is promising for implementation but remains a measurement study, not proof that substituting a simpler measure preserves treatment effectiveness. By the end of 2025, the evidence supported multicentre evaluation of effort-guided ventilation with robust outcomes, rather than routine adoption of an unconfirmed physiological target.⁹

Should REDvent routinely replace usual ventilator management in children with acute respiratory distress syndrome?

Not yet. REDvent improved a phase II weaning outcome, but multicentre confirmation is needed to establish patient benefit, safety and the contribution of effort-guided treatment.

References

1. de Vries HJ, Jonkman AH, de Grooth HJ, Duitman JW, Girbes ARJ, Ottenheijm CAC, et al. Lung- and Diaphragm-Protective Ventilation by Titrating Inspiratory Support to Diaphragm Effort: A Randomized Clinical Trial. *Crit Care Med*. 2022;50(2):192-203.
2. Hotz JC, Bornstein D, Kohler K, Smith E, Suresh A, Klein M, et al. Real-Time Effort Driven Ventilator Management: A Pilot Study. *Pediatr Crit Care Med*. 2020;21(11):933-940.
3. Khemani RG, Bhalla A, Hotz JC, Klein MJ, Kwok J, Kohler K, et al. Randomized Trial of Lung and Diaphragm Protective Ventilation in Children. *NEJM Evid*. 2025;4(6):EVIDoa2400360.
4. Khemani RG, Bhalla A, Hotz JC, Klein MJ, Kwok J, Kohler K, et al. Supplementary appendix to: Randomized Trial of Lung and Diaphragm Protective Ventilation in Children. *NEJM Evid*. 2025.
5. Inata Y, Takeuchi M. Computer Decision Support May Facilitate Liberation from Mechanical Ventilation. *NEJM Evid*. 2025;4(6):EVIDe2500082.
6. Blackwood B, Tume LN, Morris KP, Clarke M, McDowell C, Hemming K, et al. Effect of a Sedation and Ventilator Liberation Protocol vs Usual Care on Duration of Invasive Mechanical Ventilation in Pediatric Intensive Care Units: A Randomized Clinical Trial. *JAMA*. 2021;326(5):401-410.

7. Curley MA, Wypij D, Watson RS, Grant MJ, Asaro LA, Cheifetz IM, et al. Protocolized sedation vs usual care in pediatric patients mechanically ventilated for acute respiratory failure: a randomized clinical trial. *JAMA*. 2015;313(4):379-389.
8. Wong JJM, Dang H, Gan CS, Phan PH, Kurosawa H, Aoki K, et al. Lung-Protective Ventilation for Pediatric Acute Respiratory Distress Syndrome: A Nonrandomized Controlled Trial. *Crit Care Med*. 2024;52(10):1602-1611.
9. Ito Y, Herrera MG, Hotz JC, Kyogoku M, Newth CJL, Bhalla AK, et al. Estimation of inspiratory effort using airway occlusion maneuvers in ventilated children: a secondary analysis of an ongoing randomized trial testing a lung and diaphragm protective ventilation strategy. *Crit Care*. 2023;27(1):466.

TORPIDO 30/60

Oei JL, Kirby A, Travadi J, Davis P, Wright I, Ghadge A, et al. Targeted Oxygen for Initial Resuscitation of Preterm Infants: The TORPIDO 30/60 Randomized Clinical Trial. *JAMA*. 2026;335(6):523–530.

Introduction

The transition from placental gas exchange to breathing air is particularly precarious for an extremely preterm infant. The lungs may be immature and fluid-filled, respiratory effort weak and pulmonary blood flow still adapting. Supplemental oxygen can improve oxygen delivery during that transition, but immature antioxidant defences also make excessive exposure potentially harmful. The first oxygen concentration selected in the delivery room is therefore a treatment choice with competing physiological consequences.

Evidence from more mature newborns cannot simply determine the best approach in infants born several months early. Small preterm trials have produced conflicting signals. A study comparing 30% with 90% initial oxygen found less oxidative stress and less bronchopulmonary dysplasia with the lower concentration. The original TORPIDO trial, comparing room air with 100% oxygen, raised concern about higher mortality with room air in a small, non-prespecified subgroup of the most immature infants. Neither study settled the appropriate starting concentration for contemporary, saturation-guided resuscitation.^{1,2}

A 2024 individual participant data network meta-analysis added uncertainty rather than closing the question. It suggested lower mortality with initial oxygen concentrations of at least 90% compared with lower or intermediate concentrations, but the certainty of evidence was low or very low. The apparent benefit could not be assumed to apply to every concentration described as higher oxygen.³

TORPIDO 30/60 tested a more moderate contrast: starting inspired oxygen at 60% rather than 30% in infants born at 23–28 weeks' gestation, with subsequent adjustment for physiological response. It was substantially larger than the earlier individual trials and used death or brain injury at 36 weeks' corrected gestational age as its principal clinical outcome. The report first appeared online in December 2025, despite its 2026 print citation.⁴

Synopsis

TORPIDO 30/60 was a parallel-group superiority trial conducted at 31 maternity hospitals in Australia, India, Malaysia, Singapore, Spain and the United States. Recruitment ran from September 2018 to September 2024. Depending on local approval, enrolment used prospective parental consent or a waiver for randomisation and ascertainment of the primary outcome. This approach allowed participation in circumstances in which the urgency of delivery made prospective consent difficult.⁴

Eligible infants were expected to be born at 23–28 completed weeks' gestation and to require respiratory support at birth. Major cardiopulmonary or congenital abnormalities expected to affect oxygenation, neurodevelopment or survival were exclusions. Randomisation occurred shortly before delivery when birth was considered inevitable. Infants became ineligible if delivery occurred at 29 weeks or later or if the mother was transferred to a non-participating hospital before delivery.

Allocation was equal, stratified by site, gestation below or at least 26 weeks, and pregnancy multiplicity. Infants from the same multiple pregnancy received the same allocation. This avoided conflicting oxygen assignments within a delivery and required the analysis to account for correlation between siblings. Parents were masked to assignment, while delivery-room staff could see the oxygen equipment and were unmasked. The methods state that radiologists reporting neuroimaging were masked to allocation.⁴

After cord clamping, infants eligible for active care received respiratory support with an initial fraction of inspired oxygen of 0.6 or 0.3, delivered through an oxygen blender. The assigned concentration was to be maintained for at least five minutes if possible, after which oxygen could be adjusted in steps of 0.1–0.2 every 30–60 seconds towards the saturation targets. Earlier escalation to 100% oxygen was permitted for clinical concern, including persistent bradycardia or a need for advanced resuscitation. The comparison was therefore between starting strategies with rescue, rather than fixed oxygen concentrations throughout resuscitation.

The target oxygen saturation was 80–85% at five minutes and 85–95% at ten minutes and during transfer. A preductal pulse-oximetry probe was placed on the right wrist. Continuous positive airway pressure, ventilation and other aspects of delivery-room care followed local practice. Cord-clamping time was recorded, and post-admission neonatal management was left to the clinical team. These cointerventions could influence early physiological measurements and later outcomes, but they were not themselves randomised.⁴

The primary outcome was death or brain injury by 36 weeks' corrected gestational age. Brain injury was defined using neuroimaging findings, including intraventricular haemorrhage of any grade, echodense intraparenchymal lesions, periventricular leukomalacia and porencephalic cysts. The results also describe other imaging abnormalities. The endpoint was consequently broader than death or severe intraventricular haemorrhage alone. The two principal secondary outcomes were death and any brain injury considered separately.

Exploratory outcomes included early oxygen saturation and heart rate, delivery-room intubation and Apgar scores. Chest compressions, adrenaline use, early oxygen changes and specific additional brain-injury categories were identified as post hoc exploratory analyses. Several important planned outcomes, including bronchopulmonary dysplasia, retinopathy of prematurity and later survival without major disability, were not included in this report. Their absence limits assessment of the complete benefit–harm balance.⁴

The intended sample was 1,470 eligible infants, with allowance for non-adherence and clustering within multiple births. The primary analysis used a log-link model accounting for multiple births and adjusting for relevant stratification factors. No adjustment was made for multiple testing. Planned subgroup comparisons included gestation, consent approach, antenatal steroid exposure and sex. Sensitivity analyses included stillbirths and imputation of missing primary-outcome data. An independent monitoring committee reviewed interim information and recommended continuation.⁴

A total of 1,641 infants were randomised, 827 to 60% oxygen and 814 to 30%. After 99 and 73 exclusions, respectively, the retained population comprised 728 and 741 infants. Most exclusions related to delivery at 29 weeks or later or transfer or randomisation circumstances. There were also stillbirths, major abnormalities and withdrawals. Primary-outcome information was available for 703 and 720 infants, or 96.9% of the retained population. Forty-six lacked the necessary cranial-imaging information.⁴

Mean gestation was approximately 27 weeks and mean birth weight about 900 g. One-third were born before 26 weeks, approximately one-quarter were from multiple pregnancies and 54% were female. Although six countries participated, 73% were born in Australia. Almost all had some antenatal steroid exposure. These characteristics provide a broad extremely preterm cohort while leaving relatively limited information about infants at the lowest gestations or without antenatal treatment.^{4,5}

The allocated initial intervention was received by 94% of infants. Escalation to 100% oxygen was common: the main results table records 277 infants in the 60% group and 298 in the 30% group, approximately 38% and 41%. The physiological curves showed

higher oxygen exposure, saturation and heart rate with the higher starting concentration during the early minutes. Supplemental area-under-the-curve analyses supported differences over five and ten minutes, but not when extended to neonatal-unit admission. Rescue and titration progressively narrowed the exposure contrast.^{4,5}

Death or brain injury occurred in 330/703 infants assigned to 60% oxygen and 344/720 assigned to 30%: 46.9% versus 47.8%. The adjusted relative risk was 0.98, 95% confidence interval 0.89–1.09. The reported absolute risk difference was –0.84 percentage points, 95% confidence interval –6.02 to 4.35. The trial therefore did not demonstrate superiority of the higher starting concentration.⁴

Death by 36 weeks occurred in 112/728 versus 117/741 infants, 15.4% versus 15.8%, with an adjusted relative risk of 1.00, 95% confidence interval 0.79–1.27. Any brain injury was recorded in 281/662 versus 290/672, 42.4% versus 43.2%. These outcome-specific denominators differ from those used for the combined endpoint. The sensitivity analysis adding stillbirths remained neutral, as did the reported subgroup analyses.^{4,5}

Delivery-room chest compressions were recorded in 11 versus 30 infants and adrenaline use in four versus 14. These potentially favourable findings were based on small numbers and post hoc comparisons. Higher oxygen also increased the proportion with a recorded saturation of at least 95% at five minutes: 117/659 versus 86/669. Thus, the physiological effect included more high saturation readings as well as less early hypoxaemia. Only about 11% in either group were recorded within the narrow 80–85% five-minute target.⁴

Among exploratory imaging outcomes, ventriculomegaly occurred in 39/610 versus 16/620 infants, 6.4% versus 2.6%, adjusted relative risk 2.48, 95% confidence interval 1.34–4.57. Grade 3 or 4 intraventricular haemorrhage occurred in 70/697 versus 54/699, adjusted relative risk 1.30, 95% confidence interval 0.93–1.82. These findings do not establish a causal neurological hazard, but they prevent the absence of a primary-outcome difference from being treated as comprehensive reassurance about safety.⁴

Critique

TORPIDO 30/60 is a substantial advance over the small trials that previously informed initial oxygen selection in very preterm infants. It recruited a large cohort under clinically realistic delivery-room conditions and measured a consequential outcome rather than relying only on early oxygen saturation. Primary-outcome ascertainment was high within the retained population, and sibling clustering was recognised in the design and analysis. The result gives useful direct evidence about the specific choice between starting at 30% and 60% oxygen.

The neutral result does not establish equivalence. Its absolute confidence interval allows a reduction in death or brain injury of approximately six percentage points or an increase of more than four percentage points. Either could matter in this population. Mortality alone was less precisely estimated, with its interval allowing a meaningful reduction or increase. The appropriate conclusion is that superiority was not shown; the study did not demonstrate that all clinically important differences or risks have been excluded.

The treatment contrast was both pragmatic and limited. Clinicians could increase oxygen before five minutes when needed, and almost two-fifths of infants ultimately received 100% oxygen. These changes are not simply contamination: rescue was part of the assigned strategy and is inseparable from ethically acceptable delivery-room care. The trial asks whether a different starting point within that rescue-and-titration system improves outcomes. It does not compare cumulative exposure to 30% and 60% oxygen maintained over a prolonged period, or low oxygen without prompt escalation.

The physiological separation confirms that the starting concentration affected the early transition. Higher oxygen produced higher saturation and heart rate in the first minutes, so the trial was not merely a comparison of indistinguishable treatments. However, cumulative differences were no longer apparent by neonatal-unit admission. A short-lived early advantage may be too small to influence outcomes shaped by subsequent ventilation, infection, haemodynamic instability and other complications of extreme prematurity. Alternatively, beneficial and harmful effects could offset one another. The trial cannot distinguish these explanations.⁵

The oxygen-saturation target itself deserves scrutiny. Very few infants in either group were recorded within the narrow five-minute range, and higher initial oxygen increased high saturation readings as well as improving the lower end of the distribution. A strategy can reduce hypoxaemia without precisely delivering the intended target. Oxygen concentration, observed saturation and the speed of adjustment are different intervention dimensions. TORPIDO tested the first within a specified clinical framework;

it did not randomise infants to alternative saturation targets or establish the optimal titration schedule.

The relationship between early physiology and later outcome should not be assumed to be causal. Observational associations between low five-minute saturation and mortality may partly reflect underlying illness severity, poor respiratory effort or ineffective lung aeration. Increasing saturation by supplying more oxygen need not reverse all of those processes. The lack of improvement in death or brain injury despite better early measurements is a useful warning against validating a treatment solely because it moves a prognostic marker in the desired direction.⁶

The composite endpoint also requires careful naming. Death is unequivocally important, whereas imaging abnormalities vary substantially in severity and prognostic meaning. Inclusion of any-grade intraventricular haemorrhage makes the outcome broader than major neurological injury. Similar composite rates can coexist with different patterns of severe and mild injury. Longer-term developmental assessment is particularly important because cranial imaging at 36 weeks is an imperfect substitute for the child's eventual functional outcome. The neutral composite cannot be presented as proof of equivalent neurodevelopment.

The exploratory neurological findings should remain visible without becoming a competing primary conclusion. Ventriculomegaly was more common with higher oxygen, and severe intraventricular haemorrhage numerically favoured lower oxygen, but multiple outcomes were assessed and some categories were added post hoc. Small numbers, variable denominators and correlated imaging findings make selective emphasis hazardous. These observations warrant examination in subsequent follow-up and pooled analyses. They do not independently prove that starting with 60% oxygen causes brain injury.⁴

Conversely, fewer chest compressions and less adrenaline cannot establish clinical superiority after a neutral primary outcome. Those interventions were uncommon, their analyses were post hoc and treating clinicians were unmasked. Decisions to initiate advanced resuscitation can depend on the clinical response and on how staff interpret visible oxygen settings. The direction is compatible with improved early stabilisation, but replication is necessary before these findings determine a universal starting concentration. Their practical importance should be evaluated alongside potential oxygen toxicity and neurological outcomes.

The consent strategy is a notable strength with a corresponding data limitation. Waivers can improve inclusion of urgent births and avoid selecting only families with sufficient

time for a pre-delivery discussion. However, permission for secondary and exploratory data was not obtained from all families; the report states that 18% did not consent to that additional collection. Some secondary denominators therefore represent a subset of the main cohort. The available delivery-room tables should not be interpreted as complete ascertainment of every intervention in all 1,469 retained infants.⁴

Post-randomisation exclusions are understandable when allocation must precede an uncertain delivery, but they still deserve attention. The 172 excluded infants represent more than one-tenth of those randomised, and exclusions were more frequent in the higher-oxygen arm. Many could not have received the intended intervention because birth occurred outside the eligible window or setting. This differs from removing patients because of an unfavourable response after treatment. Nevertheless, the retained analysis is not literally an analysis of every randomised infant, and a sensitivity analysis adding stillbirths does not restore all other exclusions.

There are also inconsistencies in the numerical presentation of several secondary and exploratory results. The abstract reverses the group percentages for escalation to 100% oxygen relative to the main results and table. Some imaging counts, percentages and risk differences do not reconcile, and the reported five-minute saturation categories do not form a complete, internally consistent partition of their displayed denominators. The central primary-outcome figures agree across the main report and are suitable for appraisal. The less consistent secondary details warrant editorial clarification rather than reconstruction by assumption.

Generalisability is strongest for infants born at 23–28 weeks in centres able to provide blended oxygen, prompt pulse oximetry and neonatal intensive care. Most participants were Australian and antenatal steroid exposure was nearly universal. The most immature infants formed a small proportion of the cohort. Cord-clamping timing and respiratory support varied by local practice; recorded median clamping times also differed between groups. These features support pragmatic relevance but limit confidence that the same balance applies under different placental-transfusion practices or in settings with restricted equipment.

Finally, the report is an early outcome publication within a longer study. Bronchopulmonary dysplasia, retinopathy and later developmental outcomes are central to judging oxygen therapy in preterm infants. Their absence is not evidence that they were unaffected. A definitive clinical interpretation requires those outcomes alongside survival and neuroimaging, with attention to completeness of follow-up and the possibility that early physiological gains carry delayed trade-offs.

Body of Evidence

Vento and colleagues randomised 78 infants born at 24–28 weeks to initial oxygen concentrations of 30% or 90%, with subsequent saturation targeting. Lower oxygen was associated with less biochemical oxidative stress, fewer days of oxygen and ventilation, and less bronchopulmonary dysplasia. This small study provides a biological and clinical reason to avoid assuming that more oxygen is harmless. Its size, era and wider oxygen contrast limit direct comparison with TORPIDO 30/60.¹

The original TORPIDO report analysed 287 infants after allocation to room air or 100% oxygen. Recruitment stopped early after loss of equipoise about using 100% oxygen. In a non-prespecified subgroup born before 28 weeks, hospital mortality was 10/46 with room air and 3/54 with 100% oxygen. The small subgroup and early termination made this a hypothesis-generating signal. It raised a legitimate concern about insufficient initial oxygen, but did not establish the effectiveness of an intermediate 60% starting concentration.²

NETMOTION pooled individual data from 1,055 infants in 12 randomised trials and compared low, intermediate and high initial oxygen categories. High oxygen, at least 90%, was associated with lower mortality than low oxygen, with an odds ratio of 0.45 and 95% credible interval 0.23–0.86, and than intermediate oxygen, with an odds ratio of 0.34 and credible interval 0.11–0.99. Certainty was low and very low, respectively; effects on other morbidities were inconclusive. TORPIDO 30/60 directly strengthens the low-versus-intermediate comparison, but it does not test the highest category that drove the network's apparent benefit.³

The saturation references used in neonatal resuscitation also have a distinct evidential basis. Dawson and colleagues described oxygen-saturation trajectories in 468 infants who received neither oxygen nor assisted ventilation in the delivery room. These observations characterised normal transition; they did not randomise critically preterm infants to alternative targets. Reference ranges are informative, but deciding how closely an ill infant should follow them is an intervention question. This distinction helps explain why achieving a familiar saturation benchmark does not by itself prove improved outcomes.⁷

A 2025 trial by Katheria and colleagues randomised 140 infants born at 22–28 weeks to 100% or 30% oxygen during deferred cord clamping. The higher concentration increased the proportion reaching 80% saturation by five minutes, 69% versus 39%, without establishing a difference in severe intraventricular haemorrhage or death. Its exposure occurred while the cord remained intact, after which both groups followed the same

resuscitation approach. It supports an early physiological effect in a different transitional setting, while reinforcing the need for adequately powered clinical-outcome trials.⁸

By the end of 2025, TORPIDO 30/60 had narrowed an important uncertainty without identifying one optimal oxygen strategy. Its findings do not justify treating 60% oxygen as superior to 30% for preventing death or brain injury, and they do not settle the separate question of starting with 90–100%. Future work needs to integrate the starting concentration, timely titration, placental transfusion and effective ventilation, while retaining long-term survival and development as the outcomes that ultimately determine value.^{4,6}

Should resuscitation of infants born at 23–28 weeks routinely start with 60% rather than 30% oxygen to prevent death or brain injury?

TORPIDO 30/60 did not demonstrate that benefit. Higher starting oxygen improved early physiology, but the optimal strategy and its longer-term safety remain uncertain.

References

1. Vento M, Moro M, Escrig R, Arruza L, Villar G, Izquierdo I, et al. Preterm resuscitation with low oxygen causes less oxidative stress, inflammation, and chronic lung disease. *Pediatrics*. 2009;124(3):e439-49.
2. Oei JL, Saugstad OD, Lui K, Wright IM, Smyth JP, Craven P, et al. Targeted Oxygen in the Resuscitation of Preterm Infants, a Randomized Clinical Trial. *Pediatrics*. 2017;139(1):e20161452.
3. Sotiropoulos JX, Oei JL, Schmölzer GM, Libesman S, Hunter KE, Williams JG, et al. Initial Oxygen Concentration for the Resuscitation of Infants Born at Less Than 32 Weeks' Gestation: A Systematic Review and Individual Participant Data Network Meta-Analysis. *JAMA Pediatr*. 2024;178(8):774-783.
4. Oei JL, Kirby A, Travadi J, Davis P, Wright I, Ghadge A, et al. Targeted Oxygen for Initial Resuscitation of Preterm Infants: The TORPIDO 30/60 Randomized Clinical Trial. *JAMA*. 2026;335(6):523-530.
5. Oei JL, Kirby A, Travadi J, Davis P, Wright I, Ghadge A, et al. Supplemental online content to: Targeted Oxygen for Initial Resuscitation of Preterm Infants: The TORPIDO 30/60 Randomized Clinical Trial. *JAMA*. 2025.

6. Thomas AR, Foglia EE. Reconsidering Oxygen at Birth for Preterm Infants: Lessons From TORPIDO 30/60. *JAMA*. 2026;335(6):495-496.
7. Dawson JA, Kamlin CO, Vento M, Wong C, Cole TJ, Donath SM, et al. Defining the reference range for oxygen saturation for infants after birth. *Pediatrics*. 2010;125(6):e1340-7.
8. Katheria AC, Ines F, Lee HC, Sollinger C, Vali P, Morales A, et al. Deferred Cord Clamping With High Oxygen in Extremely Preterm Infants: A Randomized Clinical Trial. *JAMA Pediatr*. 2025;179(9):971-978.

Summarising, critiquing and putting into context the best critical care trials of 2025



Rob Mac Sweeney