



Critical Care Reviews

Book 2025

The Best Critical Care Trials of 2024
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 2025 brings together a selection of the major critical care trials published in 2024. Its purpose is to make important research accessible: to explain what was studied, examine how reliably each trial answers its question, and consider what the findings may mean for clinical practice.

We are delighted to share this book freely. We are grateful to the Critical Care Charitable Trust Fund of the Belfast Health and Social Care Trust for its generous sponsorship and for supporting open access to critical care education. This support helps make the work available to readers wherever they practise, teach or study.

The year produced an impressive range of trials, reflecting the commitment of patients, families, clinicians and research teams around the world. Their work advances science and strengthens the evidence on which patient care depends. Trials demonstrating benefit, identifying harm or challenging established assumptions all make valuable contributions to that progress.

Each chapter summarises the trial, appraises its strengths and limitations, and places its findings in the context of evidence available by the end of 2024. The aim is to support thoughtful discussion and informed reading of the original research.

Any selection involves judgement. Some trials have inevitably been omitted that other readers would have included, and reasonable people will interpret aspects of the evidence differently. We hope the book encourages curiosity, constructive debate and a continuing commitment to improving care for critically ill patients.

Rob Mac Sweeney
Critical Care Reviews
Belfast

Disclaimer

This book aims to summarise the major critical care trials of 2024. 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
HEMOTION	2
TRAIN	11
SAHaRA	20
ANNEXA-I	29
ENRICH	38
SWITCH	47
INTREPID	56
PROPHY-VAP	65
LASTE	74
TESLA	83
INTERACT4	92
TIMELESS	101
TRACE-III	109
■ Circulatory Trials	118
DanGer-Shock	119
PEERLESS	128
BLENDER	137
■ Resuscitation Trials	146
PARAMEDIC-3	147
IVIO	156
■ Airway Trials	165
PREOXI	166
■ Respiratory Trials	175
RENOVATE	176
HOT-COVID	185
PROLABI	193
FAST	201
VENT-AVOID	209
BIRDS	218

Table of Contents

Gastrointestinal Trials	227
REVISE	228
Nutrition Trials	237
PRECISE	238
Renal Trials	247
PROTECTION	248
REVIVAL	257
SIRAKI02	266
Haematology Trials	275
WOMAN-2	276
Sepsis Trials	285
BLING III	286
BALANCE	295
ADAPT-Sepsis	304
ASTER	313
REGARD-VAP	322
Landi-SEP	331
SCREEN	340
ACCESS	349
Trauma Trials	358
TRAUMOX2	359
Paediatric Trials	368
VODE	369
PLUSS	378
Perioperative Trials	387
OPTIMISE II	388
TIGHT-K	397
Miscellaneous Trials	406
DEFENDER	407
TELESCOPE	416
CYCLE	425



Neuro Trials

HEMOTION

Turgeon AF, Fergusson DA, Clayton L, Patton M-P, Neveu X, Walsh TS, et al. Liberal or Restrictive Transfusion Strategy in Patients with Traumatic Brain Injury. N Engl J Med 2024;391(8):722–735.

Introduction

Red-cell transfusion after traumatic brain injury sits at the intersection of two plausible but competing priorities. An injured brain may tolerate a reduction in oxygen delivery poorly, particularly when cerebral perfusion and autoregulation are impaired. Raising haemoglobin could therefore prevent secondary injury. However, increasing arterial oxygen content does not guarantee improved oxygen utilisation in injured tissue, and transfusion exposes patients to complications while consuming a finite resource. A treatment that improves a laboratory measurement must ultimately justify itself through better survival or recovery.¹

Restrictive transfusion became established in general intensive care through trials such as TRICC, which did not demonstrate a mortality advantage from a higher transfusion threshold.² Applying this practice to neurocritical care required an additional assumption: that a strategy acceptable for survival in a broad intensive care population would also preserve neurological function after brain injury. Earlier transfusion trials provided limited direct reassurance about that assumption. Neurological disability, dependence and quality of life are central outcomes in this population, and a mortality endpoint alone may miss a treatment effect that matters to survivors and their families.¹

The preceding traumatic brain injury trials were small and did not give a consistent clinical answer. One factorial trial found no neurological benefit from a higher threshold and raised concern about thrombosis; a smaller feasibility trial achieved haemoglobin separation and generated a signal favouring liberal transfusion.^{3,4} Neither established the optimal strategy for patients who had actually developed anaemia during their intensive care admission.

HEMOTION addressed that gap by enrolling adults with moderate or severe traumatic brain injury and anaemia, maintaining a clear separation between transfusion thresholds, and assessing neurological recovery six months later. Its hypothesis was that transfusing at a haemoglobin concentration of 10 g/dL or below would improve outcome compared with waiting until 7 g/dL or below.¹

Synopsis

HEMOTION was an investigator-led, international, pragmatic superiority trial conducted at 34 trauma centres with specialised neurocritical care in Canada, the United Kingdom, France and Brazil. It used an open treatment design with blinded endpoint assessment. Recruitment ran from 1 September 2017 to 13 April 2023. Funding came from the Canadian Institutes of Health Research and the Canadian Accelerating Clinical Trials Consortium. An independent monitoring committee reviewed accumulating safety information, and the investigators retained responsibility for the conduct and reporting of the trial.^{1,5}

Eligible patients were adults admitted to intensive care with moderate or severe blunt traumatic brain injury and a haemoglobin concentration of 10 g/dL or below. The injury had to be acute, with less than 24 hours between injury and the first emergency department visit; this was not a requirement to randomise within 24 hours. Moderate or severe injury was defined by a Glasgow Coma Scale score of 3–12. Important exclusions included transfusion after ICU admission but before randomisation, active life-threatening bleeding, a Glasgow Coma Scale score of 3 with bilateral fixed dilated pupils, brain death, contraindications or objection to transfusion, and a decision to withhold or withdraw life-sustaining treatment. Transfusion before ICU admission did not preclude enrolment.^{1,5}

A central, concealed, computer-generated system allocated patients equally using variable blocks of four and six, stratified by centre. The liberal strategy triggered transfusion at haemoglobin of 10 g/dL or below and the restrictive strategy at 7 g/dL or below. Both groups received leukoreduced red cells, one unit at a time, with subsequent units determined by repeat measurements. The aim was to transfuse within three hours of reaching the assigned threshold. The strategy continued until ICU discharge. Other care remained at the clinical team's discretion, with adherence to Brain Trauma Foundation guidance encouraged. Consent was obtained through surrogate or deferred arrangements, followed by patient consent when capacity returned.¹

The primary outcome was unfavourable neurological status at six months, measured using the eight-category Glasgow Outcome Scale–Extended (GOS-E). Scores range from 1, indicating death, to 8, indicating upper good recovery. A sliding dichotomy defined success relative to baseline prognosis. Patients were divided into thirds using the IMPACT prognostic model; an unfavourable outcome was GOS-E 3 or below in the worst-prognosis third, 4 or below in the intermediate third, and 5 or below in the best-prognosis third. Thus, the primary outcome did not use one fixed disability threshold for every participant. Trained research personnel, unaware of treatment allocation, assessed the main neurological and patient-reported outcomes centrally.¹

Secondary assessments included mortality, functional independence, generic and brain injury-specific quality of life, and depressive symptoms at six months. Functional independence was measured with the Functional Independence Measure (FIM), scored from 18 to 126, with higher scores indicating greater independence. The EuroQol visual analogue scale and EQ-5D-5L utility index assessed general health status; QOLIBRI assessed quality of life after brain injury. Higher scores on these instruments indicate better outcomes. The nine-item Patient Health Questionnaire assessed depressive symptoms, with higher scores indicating worse symptoms. Functional and quality-of-life analyses were conducted among survivors. Transfusion exposure, infections, complications, ventilation duration and lengths of stay were also recorded.^{1,5}

The initial calculation required 712 patients for 80% power to detect a 10-percentage-point absolute difference, at a two-sided 5% significance level. It assumed a 40% unfavourable-outcome rate in the restrictive group using GOS-E 4 or below. The report describes an increase to 742 after the interim analysis to allow for anticipated loss to follow-up. The planned halfway efficacy assessment used a stringent Haybittle–Peto boundary, and the trial continued. The primary sliding-dichotomy analysis used robust hierarchical Poisson regression, with a centre random intercept and adjustment for sex. Missing prognostic covariates were imputed. Additional analyses examined a fixed dichotomy, the full ordinal scale, adherence and missing-data assumptions. Confidence intervals for the multiple secondary comparisons were not adjusted for multiplicity.¹

There were 742 randomised patients, 371 per group. Two were subsequently identified as having aneurysmal subarachnoid haemorrhage rather than traumatic brain injury, and four withdrew consent. The reported intention-to-treat cohort therefore contained 736 patients: 369 liberal and 367 restrictive. Primary outcome data were available for 364 and 358, respectively. Mean age was 48.7 years, 72.7% were men, 73.2% had severe injury, and more than two thirds had extracranial injuries. Mean haemoglobin at randomisation was 9.1 g/dL in both groups. Median time from injury to randomisation was 55 hours in the liberal group and 56 hours in the restrictive group. Some prognostic characteristics were imbalanced, including more absent motor responses and admission hypotension in the restrictive group.¹

The intervention produced sustained separation. Median haemoglobin during intensive care was 10.8 versus 8.8 g/dL, with a reported median difference of 2.00 g/dL (95% confidence interval [CI] 1.97–2.03). At least one red-cell transfusion was given to 365/369 patients (98.9%) in the liberal group and 141/367 (38.4%) in the restrictive group. The respective totals were 1,516 and 307 units, and median units per patient were 3 and 0. Median time between measurement and transfusion was 134 versus 104 minutes.

Adjudicated protocol violations affected 14 liberal-group and seven restrictive-group patients.^{1,5}

An unfavourable primary outcome occurred in 249/364 patients (68.4%) with liberal transfusion and 263/358 (73.5%) with restrictive transfusion. The adjusted risk ratio, liberal versus restrictive, was 0.93 (95% CI 0.83–1.04). The reported adjusted absolute difference was 5.4 percentage points (95% CI –2.9 to 13.7), expressed in the opposite direction: restrictive minus liberal. The fixed GOS-E 4-or-below outcome occurred in 61.8% versus 67.0%; its unadjusted risk ratio was 0.92 (95% CI 0.83–1.03). Alternative adjusted, ordinal, complete-case and per-protocol analyses did not establish a benefit. Subgroup estimates were broadly consistent with the overall result.^{1,5}

Six-month mortality was 99/369 (26.8%) versus 96/365 (26.3%), with an adjusted hazard ratio of 1.01 (95% CI 0.76–1.35). Among survivors, the adjusted median difference in total FIM score was 4.34 points (95% CI 0.22–8.45) in favour of liberal transfusion. The corresponding differences were 5.19 points (95% CI 0.52–9.86) for the EuroQol visual analogue scale and 0.06 (95% CI 0.01–0.10) for the EQ-5D-5L utility index. QOLIBRI and depression estimates included no difference. Venous thromboembolism occurred in 31 patients in each group (8.4%). Acute respiratory distress syndrome occurred in 12/369 (3.3%) versus 3/367 (0.8%). Seven transfusion reactions were recorded among transfused patients, six in the liberal group and one in the restrictive group; none was severe. Median ICU stay was 15 days in both groups.^{1,5}

Critique

HEMOTION did not demonstrate superiority of a liberal transfusion threshold, but it did not establish that restriction was equally effective or harmless. The point estimate favoured liberal transfusion for neurological recovery, and the reported confidence interval around the absolute difference remained compatible with an important benefit. A superiority trial cannot validate non-inferiority simply because its result does not cross a significance threshold. Equally, a favourable point estimate cannot justify treating the intervention as proven. The central finding is a clinically relevant uncertainty despite a substantially better trial than the preceding evidence base.¹

Failure to deliver different treatments is not a convincing explanation for this uncertainty. Almost every liberal-group patient received blood, compared with fewer than two fifths of restrictive-group patients, and median haemoglobin differed by 2 g/dL. HEMOTION tested two recognisably different clinical strategies. It is nevertheless important to distinguish a transfusion trigger from the concentration patients actually experienced. The restrictive group did not spend its ICU stay at 7 g/dL: its median

concentration was 8.8 g/dL. Describing the trial simply as maintenance at 7 versus 10 g/dL would misrepresent both exposure and the clinical decision.¹

Patient selection sharpened the biological question but also defined its limits. Requiring anaemia ensured that the intervention could affect management. Excluding patients already transfused after ICU admission avoided including individuals whose treatment had already crossed the trial's decision point. However, that exclusion also removes an important group with early transfusion requirements. Patients with life-threatening bleeding were excluded, so the results do not address haemorrhagic resuscitation. Median randomisation roughly two days after injury places the intervention in ongoing neurocritical care rather than immediate prehospital or emergency resuscitation. No inference about withholding necessary blood during active shock follows from this study.^{1,5}

The neurological endpoint was a major strength. Six-month recovery captures consequences that ICU mortality or length of stay cannot. The sliding dichotomy also recognised that a meaningful recovery differs according to initial prognosis. Its cost is interpretative complexity: the same final GOS-E category can count as favourable in one prognostic stratum and unfavourable in another. The 73.5% restrictive-group event rate therefore should not be directly compared with the 40% planning assumption without acknowledging that the latter used a fixed GOS-E threshold. Even with the fixed definition, however, the observed restrictive-group event rate was 67.0%, indicating a substantially impaired population. The planned 10-point benefit was ambitious, and the final confidence intervals are more informative than a retrospective declaration that the trial was simply underpowered.¹

The sliding dichotomy incorporates baseline risk, but it is not identical to adjusting directly for every prognostic variable. The principal model adjusted for centre and sex, while the risk strata came from IMPACT. More restrictive-group patients had several adverse baseline features. Chance imbalance is compatible with valid randomisation and is not evidence of selection bias, but it matters when the overall estimate is modest. Analyses using additional prognostic adjustment and different representations of GOS-E did not establish superiority. Their consistency supports the broad interpretation while leaving the size and direction of any smaller treatment effect uncertain.^{1,5}

Blinded central assessment reduced the risk that knowledge of treatment influenced classification of neurological recovery. The clinical teams necessarily knew haemoglobin concentrations and transfusion allocation, leaving potential for differences in co-interventions or treatment limitation. The supplement reports broadly comparable use of many neurocritical care interventions, although individual differences occurred.

Decisions to withdraw life-sustaining treatment in ICU involved 50 patients in each group, and hospital withdrawal decisions involved 63 per group. These observations provide some reassurance but cannot exclude every effect of open treatment on subsequent care. They are preferable to treating lack of clinician blinding as an automatic reason to dismiss the trial.^{1,5}

The favourable secondary functional findings deserve attention without being promoted to confirmatory evidence. They concern outcomes important to patients, and their direction is compatible with the neurological point estimate. However, numerous scales and components were tested without multiplicity adjustment. Some confidence intervals excluded zero and others did not. These analyses also conditioned on survival, a post-randomisation event, and had more missing information than the primary endpoint. Similar overall mortality reduces one concern but does not restore a randomised comparison within the selected group of survivors. The report also notes that minimally important differences for these instruments have not been established specifically for traumatic brain injury. A four-point FIM difference cannot therefore be translated automatically into a defined clinically important gain.^{1,5}

Safety requires similarly balanced interpretation. Equal thromboembolic event rates did not reproduce the signal from the older factorial trial, but this does not prove that liberal transfusion has no thrombotic risk. The numerical excess of acute respiratory distress syndrome is relevant because respiratory complications could offset neurological benefit. Only 15 such events occurred, and the trial was not designed to estimate individual rare harms precisely. The clear practical cost was much greater blood use. A blanket liberal policy would need enough improvement in outcomes to justify that exposure and resource requirement; the primary analysis did not establish such an improvement.¹

Applicability is strongest to adults with anaemia after moderate or severe blunt injury receiving care in specialist trauma ICUs. Most participants came from high-income settings, and very few underwent invasive brain oxygen monitoring. The trial therefore informs a haemoglobin-based policy but cannot establish whether a selected patient with documented cerebral hypoxia benefits from transfusion at a different threshold. Nor does it test an intermediate trigger, such as 8 or 9 g/dL, against either extreme. The appropriate conclusion is restrained: HEMOTION leaves routine liberal transfusion unproven and provides no basis for claiming that every patient with traumatic brain injury can safely be managed to the lowest tested trigger.^{1,5}

Body of Evidence

The earlier general ICU evidence explains why a restrictive comparator was credible. TRICC randomised 838 critically ill patients after initial treatment to transfusion below 7 or below 10 g/dL, with target ranges of 7–9 and 10–12 g/dL. Thirty-day mortality was 18.7% with the restrictive strategy and 23.3% with the liberal strategy ($P=0.11$). The trial challenged routine correction of anaemia to a higher concentration. Its principal outcome and broad population, however, did not settle the neurological recovery question subsequently addressed by HEMOTION. Evidence that extra transfusion fails to improve overall ICU survival is not a direct test of whether avoiding anaemia preserves function after a specific brain injury.^{1,2}

Robertson and colleagues randomised 200 patients with closed head injury at two US trauma centres within six hours of injury. A factorial design examined erythropoietin and transfusion thresholds of 7 or 10 g/dL. Favourable six-month Glasgow Outcome Scale results occurred in 37/87 patients (42.5%) in the lower-threshold group and 31/94 (33.0%) in the higher-threshold group ($P=0.28$). Thromboembolic events were more frequent with the higher threshold: 22/101 versus 8/99. This study did not support routine maintenance above 10 g/dL, but its smaller size and early enrolment distinguish it from HEMOTION. The latter specifically required anaemia and achieved substantial haemoglobin separation, strengthening its direct test of the subsequent ICU transfusion decision.^{1,3}

The 2019 TRAHT feasibility trial took a different step towards that question. Forty-four adults with moderate or severe traumatic brain injury were assigned to thresholds of 9 or 7 g/dL for up to 14 days or ICU discharge. Mean haemoglobin was 9.3 versus 8.4 g/dL. The primary endpoint was physiological separation rather than clinical efficacy. Hospital deaths were one of 21 with liberal transfusion and seven of 23 with restriction; the neurological comparison at six months favoured liberal transfusion numerically but did not reach conventional significance. These were useful hypothesis-generating observations in a small feasibility study. They supported undertaking a larger clinical trial, rather than establishing a reliable magnitude of benefit for practice.⁴

Further direct evidence arrived during 2024. TRAIN randomised 850 adults with traumatic brain injury, subarachnoid haemorrhage or intracerebral haemorrhage and haemoglobin below 9 g/dL. It compared triggers below 9 and below 7 g/dL over 28 days. Among patients with primary outcome data, GOS-E 1–5 at 180 days occurred in 246/393 (62.6%) versus 300/413 (72.6%); the adjusted relative risk was 0.86 (95% CI 0.79–0.94). TRAIN therefore demonstrated benefit from its liberal strategy. It did not test HEMOTION's 10 g/dL trigger, was not restricted to traumatic injury, and used a fixed outcome threshold. Those differences prevent treating the trials as interchangeable, but

statistical significance in TRAIN and its absence in HEMOTION do not by themselves establish contradictory biological effects.^{1,6}

SAHARA, first published online in December 2024, contributed a further disease-specific comparison. It randomised 742 adults with aneurysmal subarachnoid haemorrhage and anaemia to mandatory transfusion at haemoglobin of 10 g/dL or below or optional transfusion at 8 g/dL or below. Unfavourable modified Rankin Scale outcome at 12 months occurred in 122/364 (33.5%) versus 136/361 (37.7%), respectively; the risk ratio was 0.88 (95% CI 0.72–1.09). Superiority was not established. Its comparator, outcome instrument, follow-up interval and disease mechanism differ from HEMOTION. In particular, the optional 8 g/dL strategy cannot be represented as another test of the restrictive 7 g/dL trigger.⁷

By the end of 2024, these studies had substantially improved the evidence base without identifying one universal neurocritical care transfusion threshold. HEMOTION supplied a large, well-separated comparison focused on anaemic patients with traumatic injury. TRAIN made a general assumption of neurological safety at the most restrictive threshold harder to sustain, while SAHARA addressed a different haemorrhagic population and comparator. The practical synthesis requires attention to the actual thresholds and enrolled patients, rather than counting trials as positive or negative. HEMOTION supports uncertainty about routinely transfusing at 10 g/dL; it should neither be used to dismiss possible benefit from less restrictive strategies nor stretched into evidence for a particular intermediate threshold that it never tested.^{1,6,7}

Should adults with traumatic brain injury and anaemia routinely be transfused at haemoglobin ≤ 10 g/dL?

HEMOTION did not establish a neurological benefit from this liberal trigger. It also did not prove that a 7 g/dL trigger is equally safe; the wider 2024 evidence and the patient's clinical circumstances remain important.

References

1. Turgeon AF, Fergusson DA, Clayton L, Patton M-P, Neveu X, Walsh TS, et al. Liberal or Restrictive Transfusion Strategy in Patients with Traumatic Brain Injury. *N Engl J Med* 2024;391(8):722–735.
2. Hébert PC, Wells G, Blajchman MA, Marshall J, Martin C, Pagliarello G, et al. A Multicenter, Randomized, Controlled Clinical Trial of Transfusion Requirements in Critical Care. *N Engl J Med* 1999;340(6):409–417.

3. Robertson CS, Hannay HJ, Yamal J-M, Gopinath S, Goodman JC, Tilley BC; Epo Severe TBI Trial Investigators. Effect of Erythropoietin and Transfusion Threshold on Neurological Recovery After Traumatic Brain Injury: A Randomized Clinical Trial. *JAMA* 2014;312(1):36–47.
4. Gobatto ALN, Link MA, Solla DJ, Bassi E, Tierno PF, Paiva W, et al. Transfusion requirements after head trauma: a randomized feasibility controlled trial. *Crit Care* 2019;23(1):89.
5. HEMOTION Trial Investigators. Supplementary Appendix to: Liberal or Restrictive Transfusion Strategy in Patients with Traumatic Brain Injury. *N Engl J Med* 2024. doi:10.1056/NEJMoa2404360.
6. Taccone FS, Rynkowski CB, Møller K, Lormans P, Quintana-Díaz M, Caricato A, et al. Restrictive vs Liberal Transfusion Strategy in Patients With Acute Brain Injury: The TRAIN Randomized Clinical Trial. *JAMA* 2024;332(19):1623–1633.
7. English SW, Delaney A, Fergusson DA, Chassé M, Turgeon AF, Lauzier F, et al. Liberal or Restrictive Transfusion Strategy in Aneurysmal Subarachnoid Hemorrhage. *N Engl J Med* 2025;392(11):1079–1088. First published online 9 December 2024.

TRAIN

Taccone FS, Rynkowski CB, Møller K, Lormans P, Quintana-Díaz M, Caricato A, et al. Restrictive vs Liberal Transfusion Strategy in Patients With Acute Brain Injury: The TRAIN Randomized Clinical Trial. JAMA 2024;332(19):1623–1633.

Introduction

A transfusion threshold developed for general intensive care may not be the best threshold for an injured brain. Anaemia reduces arterial oxygen content, while acute brain injury can impair the capacity to compensate through increased cerebral blood flow. Transfusion could therefore prevent secondary neurological injury even if it has little effect on overall survival. Conversely, raising haemoglobin does not ensure that oxygen reaches or is used by vulnerable tissue, and additional blood exposes patients to complications and consumes a limited resource.^{1,2}

The difficulty is separating these physiological arguments from evidence of clinical benefit. TRICC helped establish restrictive transfusion in general critical illness, but its mortality endpoint and population did not directly settle the question of long-term neurological recovery.³ Earlier trials after traumatic brain injury were relatively small and produced different signals. A factorial trial comparing thresholds of 7 and 10 g/dL did not show neurological benefit from the higher threshold and identified more thromboembolic events; a subsequent feasibility trial comparing 7 and 9 g/dL suggested possible benefit from liberal transfusion.^{4,5}

HEMOTION, published earlier in 2024, substantially increased the evidence base. It found a numerical reduction in unfavourable neurological outcomes with a liberal strategy, but its primary result did not establish superiority. Its confidence interval nevertheless remained compatible with meaningful benefit.⁶ Treating that finding as proof that restrictive transfusion was equally safe would have gone beyond the trial.

TRAIN examined a related but distinct question. It enrolled anaemic adults with traumatic brain injury, subarachnoid haemorrhage or intracerebral haemorrhage and compared transfusion below 9 with below 7 g/dL. By assessing neurological status at 180 days, it tested whether the additional blood use improved recovery that mattered to patients, rather than merely correcting a laboratory value.¹

Synopsis

TRAIN was an investigator-initiated, international, pragmatic, phase 3 randomised trial involving 72 ICUs in 22 countries. Treatment was open, with blinded neurological outcome assessment. Recruitment took place between 1 September 2017 and 31 December 2022, and follow-up ended on 30 June 2023. The European Society of Intensive Care Medicine and La Fondation des Geules Cassées supported the trial; the report states that funders had no role in its design, conduct, analysis or publication decisions.¹

The detailed supplementary criteria permitted patients aged 18–80 years with traumatic brain injury, non-traumatic subarachnoid haemorrhage or intracerebral haemorrhage, a Glasgow Coma Scale score of 13 or below at randomisation, and haemoglobin of 9 g/dL or below within ten days of injury. A prolonged ICU stay was expected. Both blunt and penetrating trauma were eligible, as were primary and anticoagulant-associated intracerebral haemorrhage. Major exclusions included active uncontrolled bleeding, significant pre-existing neurological disability, intracerebral haemorrhage caused by an arteriovenous malformation or tumour, inability to receive compatible blood, and a clinical requirement for haemoglobin above 9 g/dL. Patients with brain death, imminent death, a GCS of 3 with bilaterally fixed dilated pupils, or limitations on escalation of treatment were also excluded.⁷

A web-based system allocated patients in a 1:1 ratio using variable blocks of four, six and eight. Randomisation was stratified by centre, injury type and GCS category at randomisation. Allocation was concealed before assignment. Clinicians knew the assigned strategy, while patients, families and final neurological outcome assessors were reported to be unaware of it. Consent was initially obtained from legal surrogates, with consent from patients when they regained capacity.¹

Red cells were given one unit at a time when haemoglobin fell below 9 g/dL in the liberal group or below 7 g/dL in the restrictive group. The strategy continued for up to 28 days after randomisation, hospital discharge or death, whichever came first. Transfusion within a few hours was recommended, but the interval was not prescribed or recorded. Haemoglobin was measured daily according to local practice, including measurements from blood gas analysers. Other treatment and decisions about life-sustaining care remained with the clinical team. Source-document monitoring was performed for 417 of the 850 randomised patients.¹

The primary outcome was an unfavourable Glasgow Outcome Scale–Extended (GOS-E) score at 180 days. This eight-category scale runs from death at 1 to upper good recovery at 8. TRAIN classified scores 1–5 as unfavourable and 6–8 as favourable. A score of 5

represents lower moderate disability: a person may be independent in daily living but unable to resume important work or social roles. Structured telephone or face-to-face interviews with patients or relatives were conducted by healthcare professionals blinded to allocation. Secondary outcomes included 28-day mortality, the ordinal distribution of GOS-E, organ failure, a composite of death or organ failure, lengths of stay and fluid balance. Serious adverse events were collected until ICU discharge or day 28.^{1,7}

The original sample-size calculation required 2,095 patients per group for 90% power to detect a reduction in unfavourable outcome from 50% to 45%. Recruitment difficulties led to two revisions described in the report. The first increased the effect sought to ten percentage points; the second sought an eleven-point reduction with 85% power. Allowing 5% loss to follow-up, the final target was 794, with recruitment permitted to continue until the calendar deadline. The authors state that the revisions were assessed by an independent statistician and were not influenced by interim results. A single interim analysis after 300 patients used an O'Brien–Fleming approach to control overall type I error.¹

The primary comparison used complete cases, with a log-binomial model adjusted for the randomisation stratification variables. Additional analyses addressed the 14 missing neurological outcomes using imputation and opposing best-case and worst-case assignments. A post hoc analysis used GOS-E 1–4 as the unfavourable threshold. The report cautioned that secondary endpoints and adverse-event comparisons were exploratory because of multiple testing.¹

Of 8,246 people assessed, 850 were randomised: 408 to liberal and 442 to restrictive transfusion. Consent was subsequently withdrawn for 30, leaving 397 and 423 respectively in the reported study population. Primary outcome data were available for 393 and 413. Approximately 59% had traumatic injury, 23% subarachnoid haemorrhage and 18% intracerebral haemorrhage. Mean age was about 51 years, and 46% were women. Randomisation occurred a median of three days after ICU admission. Median haemoglobin was 8.5 g/dL and median GCS was 6 in both groups. About 70% underwent intracranial pressure monitoring within 48 hours of admission.¹

The strategies produced clear separation in daily haemoglobin and transfusion exposure. A total of 357/397 patients (89.9%) in the liberal group received blood during the ICU stay, compared with 205/423 (48.5%) in the restrictive group. The reported totals were 910 and 373 red-cell transfusions; median exposure was two units versus zero. Forty liberal-group patients never received blood: 21 were classified as violations, while in 19 the first haemoglobin measurement after consent had risen above 9 g/dL. Transfusions

during active bleeding in the operating theatre were not recorded in the supplementary violation table.^{1,7}

At 180 days, unfavourable neurological outcome occurred in 246/393 liberal-group patients (62.6%) and 300/413 restrictive-group patients (72.6%). The absolute difference, liberal minus restrictive, was –10.0 percentage points (95% CI –16.5 to –3.6); adjusted relative risk was 0.86 (95% CI 0.79–0.94; $P=0.002$). Missing-outcome sensitivity analyses retained the direction and statistical significance of the result. Using GOS-E 1–4, the proportions were 198/393 (50.4%) and 250/413 (60.5%), with adjusted relative risk 0.85 (95% CI 0.75–0.95). The secondary ordinal analysis also favoured liberal transfusion, with an odds ratio of 1.37 (95% CI 1.07–1.75) for better GOS-E scores.^{1,7}

There was no clear difference in 28-day mortality: 82/397 patients (20.7%) versus 94/418 (22.5%), an absolute difference of –1.8 percentage points (95% CI –7.5 to 3.8). Organ failure and lengths of stay were also similar. New cerebral ischaemia was recorded in 35/397 patients (8.8%) versus 57/423 (13.5%), relative risk 0.65 (95% CI 0.44–0.97). ARDS occurred in 29/397 (7.3%) versus 36/423 (8.5%), and venous thromboembolism in 19/397 (4.8%) versus 17/423 (4.0%). The investigators concluded that liberal transfusion reduced unfavourable neurological outcome in anaemic patients with acute brain injury.¹

Critique

TRAIN provides direct randomised evidence for a clinically important benefit from transfusing below 9 rather than below 7 g/dL in its enrolled population. The absolute primary-outcome difference is substantial, and the confidence interval excludes both no effect and very small benefit. This was supported by adjustment for the stratification variables, analyses addressing missing follow-up, an alternative functional threshold and the ordinal outcome. These findings strengthen confidence that the result is not merely an artefact of one dichotomisation. Nevertheless, the analyses use substantially the same patients and outcomes; they are checks on robustness, not independent replications.^{1,7}

The intervention was practical and clearly implemented. Haemoglobin separated, more liberal-group patients received blood, and red-cell use increased considerably. The trial therefore evaluated a real difference in treatment. It did not compare sustained haemoglobin concentrations of exactly 9 and 7 g/dL: these were transfusion triggers, and achieved concentrations were generally higher. Nor did it compare transfusion with no transfusion, since almost half the restrictive group received blood. The result supports the tested policy under trial conditions and cannot identify the minimum effective haemoglobin concentration.^{1,7}

The outcome was appropriate to the clinical problem. A patient can survive acute brain injury yet remain unable to resume a valued life, so mortality alone is an incomplete measure of success. Including lower moderate disability within the unfavourable category gives weight to participation and employment as well as basic independence. That choice also means that a ten-point reduction in the primary outcome is not synonymous with ten fewer deaths or ten fewer fully dependent survivors per hundred patients. The full ordinal analysis helps preserve the broader functional picture, although no single scale captures every patient's priorities.^{1,2}

The changes in sample size deserve explicit attention. The trial began with a much more ambitious target and ended powered for a larger treatment effect. This reduced its ability to identify smaller benefits and to give precise answers within each injury subtype. Slow recruitment and pandemic disruption are credible operational explanations, and the reported independent review and lack of dependence on interim results are reassuring. They do not make the original and final designs equivalent. The positive primary result should be assessed using its observed effect and uncertainty, while avoiding claims that the study delivered the precision originally intended.¹

Post-randomisation exclusions are a separate limitation. Thirty patients withdrew consent, including nine withdrawals attributed to attending physicians, and no subsequent outcome data were collected for them. The primary analysis therefore does not include everyone randomised. Four further liberal-group and ten restrictive-group patients lacked the 180-day outcome. Sensitivity analyses are reassuring for those 14 missing follow-ups, but do not recover outcomes for the 30 people removed after consent withdrawal. Allocation concealment protects initial assignment; it cannot by itself eliminate bias introduced by subsequent exclusion. The numbers are limited, but the distinction should remain visible.¹

Blinded outcome assessment was a major protection against biased scoring. However, clinicians knew the transfusion strategy, and associated differences in treatment intensity, rehabilitation or prognostic expectations were not fully documented. Neuroprognostication was not standardised, and withdrawal of life-sustaining treatment was not captured. Similar early mortality is reassuring but cannot exclude differential care affecting later function. The accompanying editorial emphasised this limitation while supporting the clinical importance of TRAIN's findings. Open treatment does not invalidate the result, but it makes the result an effect of an assigned care strategy, including any consequences of clinicians knowing that assignment.^{1,2}

The cerebral ischaemia result offers a plausible explanation for neurological benefit but does not establish the mechanism. Imaging was ordered at the treating clinician's

discretion, without systematic screening, and local research staff recorded new lesions. Neither lesion volume nor a standardised imaging schedule was available. Consequently, differences in investigation could alter event detection. Cerebral ischaemia was also one of many adverse-event comparisons without multiplicity correction. It is a useful supportive signal, but the primary functional outcome carries more evidential weight than this exploratory observation.^{1,7}

Safety interpretation should be similarly restrained. The trial did not identify a clear excess of ARDS or thrombosis with the liberal strategy. Individual complications were uncommon, their estimates were imprecise, and venous thromboembolism was not sought through systematic screening. These results provide reassurance without proving that additional transfusion is harmless. Blood use was demonstrably higher; the practical trade-off is whether improved neurological recovery justifies that exposure and the associated demands on local blood supply. TRAIN provides a stronger basis for accepting that trade-off than a physiological rationale alone.^{1,2}

The multinational population broadens relevance, but substantial differences remain between participating settings. All centres listed in the appendix had an on-site physician around the clock, and enrolment was concentrated in a minority of hospitals. The trial includes patients from differing income settings, but does not establish that a liberal policy can be implemented equally where compatible blood or transfusion monitoring is limited. Recruitment over more than five years also encompassed changes in clinical care and pandemic conditions. These do not negate randomisation, but matter when translating the strategy into a service with different resources.^{1,7}

The clearest application is to anaemic adults with the studied acute injuries, persistent impairment of consciousness and a substantial anticipated ICU stay. Most had traumatic injury; the smaller haemorrhage subgroups do not independently establish equal benefit for every diagnosis. A lack of a statistically significant interaction is not proof that effects are identical. Patients over 80, those with major pre-existing disability, active uncontrolled bleeding, ischaemic stroke or post-anoxic coma were outside the detailed eligibility criteria. TRAIN supports a 9 g/dL trigger over 7 g/dL in the population it studied; it does not justify routine transfusion to higher untested targets or substitute for resuscitation during major haemorrhage.^{1,7}

Body of Evidence

TRICC established the importance of questioning routine liberal transfusion in critical illness. In 838 patients, it compared triggers below 7 and below 10 g/dL, with target ranges of 7–9 and 10–12 g/dL. Thirty-day mortality was 18.7% and 23.3%, respectively, without a statistically significant difference ($P=0.11$). Its broader ICU population and survival endpoint supported restraint in blood use but did not provide a direct assessment of neurological recovery after acute brain injury. TRAIN addresses this gap rather than overturning every indication for restrictive transfusion.^{1,3}

The 2014 Robertson trial randomised 200 patients with closed head injury within six hours in a factorial comparison of erythropoietin and transfusion thresholds of 7 or 10 g/dL. Favourable six-month Glasgow Outcome Scale results occurred in 37/87 patients (42.5%) with the lower threshold and 31/94 (33.0%) with the higher threshold ($P=0.28$). Thromboembolic events were more frequent with the higher threshold, 22/101 versus 8/99. This earlier result argues against assuming that any higher threshold must be beneficial. Its early recruitment, factorial design and different liberal trigger distinguish it from TRAIN's comparison after anaemia had developed.^{1,4}

The 2019 TRAHT feasibility trial offered a closer comparison, assigning 44 adults with moderate or severe traumatic brain injury to thresholds of 9 or 7 g/dL. It achieved mean haemoglobin concentrations of 9.3 and 8.4 g/dL, respectively. Hospital deaths numbered one of 21 with liberal and seven of 23 with restrictive transfusion; the six-month neurological comparison favoured liberal transfusion numerically but did not reach conventional significance. Its primary purpose was to establish feasibility through haemoglobin separation. TRAIN converted that promising but small clinical signal into a substantially larger multicentre test.⁵

HEMOTION randomised 742 anaemic adults with moderate or severe traumatic brain injury to triggers of 10 or 7 g/dL. Its primary outcome used a prognosis-dependent sliding GOS-E dichotomy at six months, rather than TRAIN's fixed 1–5 definition. Unfavourable outcomes occurred in 68.4% and 73.5%; the adjusted difference, restrictive minus liberal, was 5.4 percentage points (95% CI –2.9 to 13.7). HEMOTION did not establish superiority, but its interval includes benefit of the magnitude observed in TRAIN. Different significance labels are insufficient to conclude that the studies conflict. The populations, liberal triggers and outcome definitions also prevent treating them as identical experiments.^{1,6}

SAHARA supplied a further 2024 comparison, first published online in December. It randomised 742 adults with aneurysmal subarachnoid haemorrhage and anaemia to mandatory transfusion at haemoglobin of 10 g/dL or below or optional transfusion at 8

g/dL or below. At 12 months, modified Rankin Scale scores of 4–6 occurred in 122/364 patients (33.5%) and 136/361 (37.7%), relative risk 0.88 (95% CI 0.72–1.09; P=0.22). This did not establish superiority. Because its comparator allowed transfusion at 8 g/dL, it cannot be read as another demonstration of the safety or otherwise of waiting until 7 g/dL. It also assessed a different population, disability scale and follow-up interval.⁸

Taken together, the evidence available by the end of 2024 makes a blanket assumption that a 7 g/dL threshold is appropriate for every neurocritically ill patient difficult to defend. TRAIN provides the clearest positive randomised comparison of 9 against 7 g/dL in anaemic acute brain injury. HEMOTION and SAHARA constrain how far that finding should be extended to different triggers and diagnoses. The editorial accompanying TRAIN favoured liberal transfusion; that interpretation is persuasive for patients resembling those enrolled, while a single optimal target across neurocritical care remains unestablished. Future refinement may involve injury-specific risk or cerebral oxygenation, but TRAIN itself tested a haemoglobin-based policy.^{1,2,6,8}

Should anaemic adults with acute brain injury be transfused below 9 rather than below 7 g/dL?

For patients resembling those enrolled in TRAIN, the 9 g/dL trigger improved 180-day neurological outcome and is supported over the 7 g/dL trigger. The trial does not establish a higher target or equal benefit in every brain-injury subgroup.

References

1. Taccone FS, Rynkowski CB, Møller K, Lormans P, Quintana-Díaz M, Caricato A, et al. Restrictive vs Liberal Transfusion Strategy in Patients With Acute Brain Injury: The TRAIN Randomized Clinical Trial. *JAMA* 2024;332(19):1623–1633.
2. Turgeon AF, Lauzier F. Shifting Balance of the Risk-Benefit of Restrictive Transfusion Strategies in Neurocritically Ill Patients—Is Less Still More? *JAMA* 2024;332(19):1615–1617.
3. Hébert PC, Wells G, Blajchman MA, Marshall J, Martin C, Pagliarello G, et al. A Multicenter, Randomized, Controlled Clinical Trial of Transfusion Requirements in Critical Care. *N Engl J Med* 1999;340(6):409–417.

4. Robertson CS, Hannay HJ, Yamal J-M, Gopinath S, Goodman JC, Tilley BC; Epo Severe TBI Trial Investigators. Effect of Erythropoietin and Transfusion Threshold on Neurological Recovery After Traumatic Brain Injury: A Randomized Clinical Trial. *JAMA* 2014;312(1):36–47.
5. Gobatto ALN, Link MA, Solla DJ, Bassi E, Tierno PF, Paiva W, et al. Transfusion requirements after head trauma: a randomized feasibility controlled trial. *Crit Care* 2019;23(1):89.
6. Turgeon AF, Fergusson DA, Clayton L, Patton M-P, Neveu X, Walsh TS, et al. Liberal or Restrictive Transfusion Strategy in Patients with Traumatic Brain Injury. *N Engl J Med* 2024;391(8):722–735.
7. TRAIN Study Group. Supplemental Online Content to: Restrictive vs Liberal Transfusion Strategy in Patients With Acute Brain Injury: The TRAIN Randomized Clinical Trial. *JAMA* 2024. doi:10.1001/jama.2024.20424.
8. English SW, Delaney A, Fergusson DA, Chassé M, Turgeon AF, Lauzier F, et al. Liberal or Restrictive Transfusion Strategy in Aneurysmal Subarachnoid Hemorrhage. *N Engl J Med* 2025;392(11):1079–1088. First published online 9 December 2024.

Possible Reporting Inconsistencies. The TRAIN discussion describes HEMOTION’s unfavourable outcome as GOS-E 1–4. HEMOTION’s primary analysis actually used a prognosis-dependent sliding dichotomy; GOS-E 1–4 was an additional fixed-threshold analysis. This distinction matters when comparing the two trials.^{1,6}

English SW, Delaney A, Fergusson DA, Chassé M, Turgeon AF, Lauzier F, et al. Liberal or Restrictive Transfusion Strategy in Aneurysmal Subarachnoid Hemorrhage. N Engl J Med 2025;392(11):1079–1088. First published online 9 December 2024.

Introduction

The consequences of aneurysmal subarachnoid haemorrhage extend beyond the initial bleed. Delayed cerebral ischaemia and other secondary insults can worsen recovery after an aneurysm has been secured. Anaemia is common during this period and may reduce oxygen delivery to tissue already threatened by impaired perfusion. A liberal red-cell transfusion strategy is therefore biologically plausible, but increasing circulating haemoglobin is an intermediate goal. The clinical question is whether the extra transfusions leave more people alive with an acceptable level of neurological function.¹

Observational associations are particularly difficult to interpret in this setting. Lower haemoglobin can mark greater illness severity, procedural blood loss or a complicated ICU course. Transfusion is itself more likely in the sickest patients. Associations between anaemia, blood administration and poor outcome cannot reliably distinguish the effects of treatment from the reasons for giving it. Randomised comparisons are needed, with follow-up long enough to capture recovery beyond hospital discharge.¹

Earlier randomised evidence specific to subarachnoid haemorrhage was limited. A small pilot study showed that higher haemoglobin goals could be achieved, but it compared two relatively high targets and could not establish a durable clinical advantage.² Broader neurocritical care evidence expanded markedly during 2024. HEMOTION did not establish a significant primary benefit from liberal transfusion after traumatic brain injury, whereas TRAIN found better neurological outcomes with a 9 rather than 7 g/dL trigger in a mixed acute brain injury population.^{3,4} Neither directly answered the question of routinely transfusing at 10 g/dL after aneurysm rupture compared with a strategy permitting blood at 8 g/dL.

SAHARA addressed that disease-specific decision. It compared mandatory transfusion at haemoglobin of 10 g/dL or below with optional transfusion at 8 g/dL or below, and assessed neurological function a year later. Its December 2024 online publication completed an important year of transfusion research in neurocritical care.¹

Synopsis

SAHARA was an investigator-initiated, pragmatic, open-label superiority trial with blinded outcome assessment. It was conducted at 23 specialist neurocritical care centres with neurosurgical and neurointerventional support in Canada, Australia and the United States. Public research funding supported the study, and the report states that funders had no role in design, analysis, interpretation or manuscript preparation. A central coordinating team and an independent data and safety monitoring board oversaw the trial. Consent before enrolment or delayed consent to continue was used according to local approvals.¹

An internal pilot contributed 60 participants enrolled between 17 October 2015 and 21 November 2016. The subsequent full trial enrolled 682 between 11 March 2018 and 10 July 2023. Pilot outcomes were incorporated into the final analysis. Adults with a first-ever aneurysmal subarachnoid haemorrhage and haemoglobin of 10 g/dL or below within ten days of hospital presentation were eligible; the pilot had allowed fourteen days. Exclusions included non-aneurysmal haemorrhage, active bleeding with haemodynamic instability, a contraindication or objection to transfusion, and a decision to withhold or withdraw active medical care. The supplementary criteria also excluded haemorrhage due to a mycotic aneurysm, infundibulum or vascular malformation.^{1,5}

Patients were assigned in a 1:1 ratio through a central, concealed, computer-generated sequence, using variable blocks of four and six and stratification by centre. The liberal group received mandatory transfusion when haemoglobin was 10 g/dL or below. In the restrictive group, transfusion was permitted only at 8 g/dL or below and remained optional. Both groups received leukoreduced red cells one unit at a time. The assigned strategy continued from randomisation until day 21 after the first hospital presentation, hospital discharge or death, whichever occurred first. Thus, it was not a fixed 21-day course beginning at randomisation.¹

Haemoglobin was measured by the local laboratory or point-of-care testing, daily through hospital day ten or while the patient remained in intensive care. Other management was determined by treating clinicians, with adherence to relevant clinical guidance encouraged. The trial recorded threshold adherence, time to transfusion, blood exposure and cointerventions. Central assessors who were unaware of allocation obtained the long-term outcomes from patients or their proxies.¹

The primary outcome was unfavourable neurological function at 12 months, defined as a modified Rankin Scale (mRS) score of 4–6. The scale runs from 0, no symptoms, to 6, death; a score of 4 indicates moderately severe disability requiring assistance with walking and bodily needs. This endpoint therefore included death and substantial

dependency. Secondary assessments at 12 months included the Functional Independence Measure (FIM), EQ-5D-5L utility index and a visual analogue scale of perceived health. Higher scores on these measures indicate better function or health. The investigators reported analyses including nonsurvivors and additional analyses restricted to survivors.^{1,5}

Early secondary outcomes included new radiographic vasospasm, delayed cerebral ischaemia and cerebral infarction. The latter was captured through routinely obtained brain imaging. An independent assessor, blinded to allocation, centrally adjudicated these outcomes. The appendix describes delayed cerebral ischaemia as an otherwise unexplained neurological deterioration lasting at least an hour, including a fall in GCS of at least two points or a new focal deficit. This clinical outcome was distinguished from vessel narrowing on imaging and from new infarction. Further outcomes included haemoglobin concentrations, transfusion exposure, ventilation, lengths of stay and adverse events.^{1,5}

A sample of 740 was calculated to provide 80% power, with a two-sided alpha of 0.05 and allowance for 3% loss to follow-up, to detect a 25% relative reduction from an assumed 40% unfavourable-outcome rate. The primary model used a log-link binomial mixed model with treatment and sex as fixed effects and centre as a random effect. The full ordinal mRS was also examined with a proportional-odds model. A halfway analysis after 370 patients used a Haybittle–Peto stopping rule. Sensitivity analyses examined additional adjustment for age and clinical severity, an mRS threshold of 3, adherence and missing-data assumptions. Secondary and subgroup confidence intervals were not adjusted for multiple comparisons.¹

Of 742 randomised patients, 372 entered the liberal group and 370 the restrictive group. Nine subsequently declined or withdrew consent; data for another patient who died before delayed consent could be obtained were excluded at the local ethics board's direction. This left 366 per group in the study population. The 12-month primary outcome was available for 364 liberal-group and 361 restrictive-group patients, or 725/742 of those randomised. Seven patients enrolled on hospital day eleven, outside the full trial's window, remained in the main analysis.^{1,5}

Mean age was 59.4 years, and 82% were women. Three quarters were enrolled in Canada. Randomisation occurred a median of three days after hospital presentation, with median haemoglobin of 9.5 g/dL in both groups. Median admission World Federation of Neurological Surgeons grade was 3, and median modified Fisher grade was 4. Around 70% had an external ventricular drain. Most aneurysms were in the anterior circulation and were secured promptly, with endovascular treatment the more common

approach. Previous transfusion had been given to 40/366 liberal-group and 28/366 restrictive-group patients.^{1,5}

The intervention lasted a median of 17 and 16 days, respectively. Haemoglobin concentrations separated throughout the intervention period. Median pretransfusion haemoglobin was 9.6 g/dL with liberal and 7.6 g/dL with restrictive treatment; median time from measurement to transfusion was 6.1 and 4.4 hours. Red cells were administered to 365/366 patients (99.7%) versus 128/366 (about 35%), with median exposure of two units versus zero. The reported median difference was 2.0 units (95% CI 1.78–2.20). Twelve major threshold or timing violations affected eleven patients, while lesser departures, particularly delayed transfusion in the liberal group, were more frequent.^{1,5}

Unfavourable mRS outcome occurred in 122/364 patients (33.5%) in the liberal group and 136/361 (37.7%) in the restrictive group. The adjusted risk ratio was 0.88 (95% CI 0.72–1.09; P=0.22). The ordinal odds ratio for a worse mRS score was 0.82 (95% CI 0.64–1.06). Sensitivity analyses did not change the overall interpretation. Using mRS 3–6, unfavourable outcomes occurred in 178/364 (48.9%) and 197/361 (54.6%), with risk ratio 0.89 (95% CI 0.77–1.04).^{1,5}

Mortality at 12 months was 99/364 (27.2%) versus 98/361 (27.1%), risk ratio 0.99 (95% CI 0.77–1.28). Functional and quality-of-life comparisons did not show a clear difference. Among survivors, mean FIM was 114.1 versus 110.2, with adjusted mean difference 3.84 (95% CI –0.78 to 8.47). New radiographic vasospasm was recorded in 115/365 patients (31.5%) versus 149/366 (40.7%), risk ratio 0.77 (95% CI 0.63–0.95). Delayed cerebral ischaemia occurred in 64/365 (17.5%) versus 73/366 (19.9%), and new infarction in 58/365 (15.9%) versus 65/366 (17.8%), without clear differences. ICU and hospital stays were similar. Acute transfusion reactions were reported in two liberal-group and one restrictive-group patients; no unexpected adverse events occurred. The investigators concluded that liberal transfusion did not reduce the risk of unfavourable neurological outcome.^{1,5}

Critique

SAHARA supplies a substantial randomised comparison in the population that matters to the question. Allocation concealment, a clear intervention contrast, central blinded functional assessment and high follow-up completeness are important strengths. The principal result did not establish superiority of mandatory transfusion at 10 g/dL over optional transfusion at 8 g/dL. That is the central finding. It does not establish equivalence: the confidence interval remains compatible with a meaningful relative benefit as well as modest harm. Describing the strategies as interchangeable would remove uncertainty that the trial has not resolved.¹

The distinction between mandatory and optional treatment is essential. SAHARA evaluated two policies, not merely two laboratory values. Almost every liberal-group patient received blood, whereas approximately two thirds of the restrictive group avoided it. The restrictive policy therefore allowed clinicians to individualise transfusion once haemoglobin reached 8 g/dL. It should not be rewritten as compulsory transfusion at that concentration, nor as a strategy that required waiting until 7 g/dL. This difference is especially important when comparing SAHARA with TRAIN, which tested a lower restrictive trigger.^{1,4}

The biological contrast was achieved. Pretransfusion haemoglobin differed by about 2 g/dL and serial measurements showed sustained separation. A failure to implement the assigned strategies is therefore an unconvincing general explanation for the absence of a significant primary benefit. Nevertheless, the low reported number of major violations does not imply immediate, uniform treatment. The appendix distinguishes major violations from lesser deviations, including transfusion six to 24 hours after the trigger was reached. The result reflects a pragmatic hospital policy with those implementation characteristics, not an instantaneous correction of every low haemoglobin measurement.^{1,5}

Precision remains a major limitation. The study was designed to detect a relatively large improvement: a ten-percentage-point reduction implied by a 25% relative reduction from a 40% control risk. The observed restrictive-group event rate was reasonably close to that assumption, but the observed difference was smaller. The confidence interval still includes the relative effect used in the sample-size calculation. Consequently, the trial's negative significance test does not exclude all clinically worthwhile benefit. This matters because even a modest improvement in long-term independence could be valuable after a disabling brain injury.¹

A year of follow-up is a strength in a condition with prolonged recovery. The primary mRS threshold also captures severe outcomes that patients and families understand.

However, combining death with mRS 4 and 5 and treating scores 0–3 as favourable compresses a wide spectrum of recovery. A person who walks independently but needs other help is counted with someone who has no symptoms. The ordinal analysis examines more of that information and likewise failed to demonstrate a clear benefit, but its interval does not prove that smaller shifts in function are absent. Fatigue, cognition, mood and ability to resume valued roles may also escape a global disability score.^{1,5}

The additional functional and quality-of-life measures are therefore useful. Their results point in broadly the same direction as the primary estimate without establishing a clear difference. Interpretation requires care about whether deaths are included and how they are scored; analyses limited to survivors answer a different question from an outcome assessed across the randomised population. Survivor comparisons can also lose the protection of randomisation through conditioning on survival. Nearly identical mortality reduces one concern but does not remove that principle. Minimal important differences for these scales have not been established specifically for subarachnoid haemorrhage, further limiting claims about the clinical meaning of small score differences.^{1,5}

Missing primary outcomes were uncommon. Seven participants with retained consent lacked a 12-month score, and sensitivity analyses using opposing assumptions produced the same broad conclusion. Ten additional randomised patients were removed for consent-related reasons, including one known death whose data could not be retained. The reasons are transparently reported, and the numbers are small, but the main comparison is not a complete analysis of every randomised participant. Likewise, the per-protocol analysis supports robustness without replacing the primary policy comparison; removing or reweighting participants for adherence relies on additional assumptions.^{1,5}

The reduction in radiographic vasospasm is interesting but should not be promoted to proof of improved cerebral perfusion or patient benefit. Vessel narrowing, clinical delayed ischaemia and completed infarction are related but distinct outcomes. Their ascertainment is challenging, and the central adjudication process depended on clinical and imaging information generated by the treating centres. Blinding the adjudicator helps classification but cannot fully standardise which investigations were obtained. The vasospasm confidence interval also arose among multiple secondary comparisons. Its failure to coincide with clear reductions in delayed cerebral ischaemia, infarction or disability argues for a cautious, exploratory interpretation.^{1,5}

Clinician blinding was impractical, leaving the possibility that knowledge of assignment altered other care. The appendix provides useful information about blood pressure

targets, fluids, vasopressors and treatments for vasospasm or delayed ischaemia. Some interventions were numerically more common in the restrictive group. These differences may reflect a response to the assigned strategy or evolving illness, rather than baseline confounding. They should not automatically be adjusted away, because the clinically relevant effect is that of adopting the transfusion policy within routine care. They do, however, limit any claim that haemoglobin alone explains the observed outcomes.^{1,5}

No major safety signal emerged, but individual transfusion complications were too uncommon to establish equal safety. The exposure difference is certain; a reduction in disability is not. This balance gives clinicians a reason to avoid treating haemoglobin of 10 g/dL as an obligatory transfusion target for every patient resembling those enrolled, while preserving discretion for particular physiological concerns. Generalisation is strongest to adults with anaemia after a first aneurysmal haemorrhage managed in specialist centres. All sites were in high-income countries, women were strongly represented, and active unstable bleeding was excluded. SAHARA does not answer emergency haemorrhage resuscitation, non-aneurysmal subarachnoid haemorrhage, or transfusion guided by documented brain tissue hypoxia.^{1,5}

Body of Evidence

The broader history of transfusion trials explains why liberal correction of anaemia needed a direct test. TRICC randomised 838 critically ill patients to restrictive or liberal policies, using transfusion triggers below 7 or 10 g/dL. Thirty-day mortality was 18.7% and 23.3%, respectively, without a statistically significant difference. This challenged routine liberal transfusion in general intensive care, but did not resolve a disease-specific question about long-term function after aneurysm rupture. Applying its result to every neurological injury would require evidence beyond its population and endpoint.⁶

Naidech and colleagues provided an earlier randomised study directly in subarachnoid haemorrhage. Forty-four patients at high risk of vasospasm were assigned to haemoglobin goals of at least 10 or 11.5 g/dL. The higher-goal group received more blood, and haemoglobin separated from day four. Cerebral infarction on MRI occurred in six of 20 versus nine of 22 patients, while clinical outcomes favoured the higher goal numerically without statistically significant differences. This was a pilot feasibility and safety comparison, too small to establish clinical efficacy. Both targets were also above SAHARA's restrictive threshold. It supported further investigation, not routine maintenance of very high haemoglobin.²

HEMOTION, published in June 2024, examined adults with moderate or severe traumatic brain injury and anaemia. It randomised 742 patients to triggers of 10 or 7 g/dL and used

a prognosis-dependent sliding GOS-E dichotomy at six months. Unfavourable outcomes occurred in 68.4% and 73.5%, respectively; the adjusted difference, restrictive minus liberal, was 5.4 percentage points (95% CI -2.9 to 13.7). It did not establish superiority, although selected exploratory functional measures favoured liberal transfusion. HEMOTION and SAHARA therefore both leave residual uncertainty about neurological benefit from a 10 g/dL trigger, while studying different injuries, comparators and outcomes.³

TRAIN, published in October 2024, randomised 850 adults with traumatic brain injury, subarachnoid haemorrhage or intracerebral haemorrhage and anaemia. It compared triggers below 9 and below 7 g/dL for up to 28 days. GOS-E 1–5 at 180 days occurred in 246/393 patients (62.6%) with liberal and 300/413 (72.6%) with restrictive treatment, adjusted relative risk 0.86 (95% CI 0.79–0.94). The study population included 190 patients with subarachnoid haemorrhage. The injury-type subgroup analysis did not identify clear heterogeneity, but that subgroup was much smaller than SAHARA's dedicated population. TRAIN provides direct support for its 9 versus 7 g/dL policy comparison, not for mandatory transfusion at 10 versus optional transfusion at 8 g/dL.⁴

The three large 2024 trials should therefore be compared by the questions they asked, rather than counted as positive or negative votes. All primary point estimates favoured the liberal policy, but only TRAIN established a statistically significant primary benefit. Different thresholds, prognostic selection, disability definitions and follow-up intervals can alter both event rates and treatment effects. In particular, SAHARA cannot demonstrate that a 7 g/dL trigger is safe after aneurysm rupture because it never assigned that strategy. Equally, TRAIN cannot prove that every patient with subarachnoid haemorrhage benefits from mandatory transfusion at 10 g/dL.^{1,3,4}

By the end of 2024, SAHARA had replaced a major area of extrapolation with a large disease-specific randomised comparison. Its practical contribution is to show that a substantially greater use of blood at the 10 g/dL trigger did not produce an established improvement in 12-month disability over a policy permitting transfusion at 8 g/dL. The remaining question is whether a smaller average benefit, a different threshold or a more selective physiological indication can justify additional transfusion. Those possibilities remain open; they are not demonstrated by the exploratory vasospasm finding or excluded by the primary P value.¹

Should all anaemic patients with aneurysmal subarachnoid haemorrhage receive mandatory transfusion at haemoglobin ≤ 10 g/dL?

SAHARA did not establish better 12-month neurological outcomes than a strategy allowing optional transfusion at ≤ 8 g/dL. It leaves meaningful uncertainty about smaller benefits and does not test a 7 g/dL trigger.

References

1. English SW, Delaney A, Fergusson DA, Chassé M, Turgeon AF, Lauzier F, et al. Liberal or Restrictive Transfusion Strategy in Aneurysmal Subarachnoid Hemorrhage. *N Engl J Med* 2025;392(11):1079–1088. First published online 9 December 2024.
2. Naidech AM, Shaibani A, Garg RK, Duran IM, Liebling SM, Bassin SL, et al. Prospective, randomized trial of higher goal hemoglobin after subarachnoid hemorrhage. *Neurocrit Care* 2010;13(3):313–320.
3. Turgeon AF, Fergusson DA, Clayton L, Patton M-P, Neveu X, Walsh TS, et al. Liberal or Restrictive Transfusion Strategy in Patients with Traumatic Brain Injury. *N Engl J Med* 2024;391(8):722–735.
4. Taccone FS, Rynkowski CB, Møller K, Lormans P, Quintana-Díaz M, Caricato A, et al. Restrictive vs Liberal Transfusion Strategy in Patients With Acute Brain Injury: The TRAIN Randomized Clinical Trial. *JAMA* 2024;332(19):1623–1633.
5. SAHARA Trial Investigators. Supplementary Appendix to: Liberal or Restrictive Transfusion Strategy in Aneurysmal Subarachnoid Hemorrhage. *N Engl J Med*. doi:10.1056/NEJMoa2410962.
6. Hébert PC, Wells G, Blajchman MA, Marshall J, Martin C, Pagliarello G, et al. A Multicenter, Randomized, Controlled Clinical Trial of Transfusion Requirements in Critical Care. *N Engl J Med* 1999;340(6):409–417.

ANNEXA-I

Connolly SJ, Sharma M, Cohen AT, Demchuk AM, Członkowska A, Lindgren AG, et al. Andexanet for Factor Xa Inhibitor–Associated Acute Intracerebral Hemorrhage. *N Engl J Med* 2024;390(19):1745–1755.

Introduction

An intracerebral haemorrhage during factor Xa inhibitor treatment creates two competing emergencies. Continued bleeding may enlarge the haematoma and cause irreversible neurological injury, while reversing anticoagulation exposes a patient with an underlying thrombotic indication to renewed ischaemic risk. A treatment that rapidly corrects an anticoagulation assay therefore needs to demonstrate more than biochemical activity. Its value depends on preventing clinically consequential bleeding without replacing haemorrhagic injury with thrombotic injury. This distinction is particularly important when the patient is older, has atrial fibrillation and may already have substantial cerebrovascular disease.¹

Andexanet alfa is a modified, catalytically inactive factor Xa protein that binds factor Xa inhibitors. Earlier randomised studies in healthy volunteers demonstrated rapid reversal of apixaban and rivaroxaban activity. The subsequent ANNEXA-4 cohort reported haemostatic efficacy in patients with major bleeding, but lacked a concurrent control group. Consequently, neither the clinical benefit attributable to andexanet nor its excess thrombotic risk could be reliably separated from the natural history of severe bleeding and interruption of anticoagulation.^{2,3}

Prothrombin complex concentrate was commonly used as an alternative, although uncertainty remained about the optimal management of factor Xa inhibitor associated intracerebral haemorrhage. The decisive comparison was therefore between specific reversal and contemporary usual care, usually including concentrate. ANNEXA-I supplied that randomised comparison. Its primary question concerned haemostatic efficacy during the first 12 hours, with thrombotic events and death assessed over 30 days.¹

The trial established a measurable effect on early bleeding control. Whether that effect justified routine treatment was more difficult. The same study found more thrombotic events, particularly ischaemic stroke, without establishing better functional recovery or survival. Interpreting ANNEXA-I requires keeping these different outcomes, their time horizons and the populations used for their analyses in view throughout.¹

Synopsis

ANNEXA-I was an international, randomised, open label trial with blinded outcome assessment. Recruitment took place at 131 sites in 23 countries between June 2019 and May 2023. Portola Pharmaceuticals, subsequently acquired by Alexion AstraZeneca Rare Disease, sponsored the trial. The Population Health Research Institute coordinated the study and performed the statistical analysis. An independent data and safety monitoring board reviewed accumulating data.¹

Eligibility changed during recruitment. Initially, the study included several forms of acute intracranial haemorrhage during factor Xa inhibitor treatment. Later amendments restricted recruitment to intracerebral haemorrhage with a volume of 0.5–60 mL and a National Institutes of Health Stroke Scale (NIHSS) score no greater than 35. The permitted interval from symptom onset to baseline imaging was shortened from 12 to six hours. Baseline imaging had to precede randomisation by no more than two hours. The principal report specified factor Xa inhibitor use within 15 hours before randomisation; the appendix described an alternative laboratory criterion when the last dose was earlier or its timing unknown.^{1,4}

Patients with a Glasgow Coma Scale score below seven, a planned operation within 12 hours or a thrombotic event within the preceding two weeks were excluded. These criteria selected patients in whom early haemostatic treatment could be assessed before immediate neurosurgery or profound neurological deterioration overtook the intervention. They also limit extrapolation to very large haemorrhages, deeply unconscious patients and those requiring urgent evacuation.¹

Participants were allocated equally to andexanet or usual care. Following an amendment, randomisation was stratified by intended prothrombin complex concentrate use and the interval from symptom onset to baseline imaging, divided at 180 minutes. Andexanet was administered as an intravenous bolus over 15–30 minutes followed by a two-hour infusion. A low or high dose regimen was selected according to the anticoagulant, its dose and the interval since administration. Usual care was chosen by the treating clinician and could include concentrate, but excluded planned andexanet administration.^{1,4}

Treating teams knew allocation. Central imaging readers, the assessor performing the 12-hour NIHSS examination and the committee adjudicating efficacy and thrombotic events were blinded. The primary endpoint required all three of the following: haematoma enlargement no greater than 35% at 12 hours, an NIHSS deterioration of less than seven points at 12 hours, and no rescue treatment between three and 12 hours. Rescue treatment included additional andexanet, concentrate or decompressive surgery.

Missing assessments were adjudicated, and patients without an assessable primary outcome were counted as having poor or no haemostatic efficacy.¹

The principal secondary endpoint was the percentage reduction in anti-factor Xa activity from baseline to its nadir during the first two hours. Safety outcomes included thrombotic events and death within 30 days. The modified Rankin scale at 30 days was an exploratory functional measure; the reported dichotomous analysis was post hoc. The study was designed around haemostatic efficacy, rather than a definitive test of disability or mortality.¹

The planned sample increased from 440 to 900 patients. The revised calculation provided 90% power to detect a ten percentage point improvement in haemostatic efficacy, assuming approximately 70% efficacy with usual care. An interim analysis was planned after approximately 450 patients. The report states that its stopping boundary was changed from $P < 0.001$ to $P < 0.031$ before knowledge of the treatment effect. Primary and key secondary testing followed a hierarchy intended to control the overall type I error rate.¹

Of 550 randomised patients, 20 had consent irregularities requiring removal of their data, leaving 530: 263 assigned to andexanet and 267 to usual care. The interim primary analysis included 452 patients, 224 and 228 respectively. Another 78 entered before recruitment stopped following the monitoring board's review. Thus, the main efficacy result and the principal safety results used different, explicitly described populations.¹

In the efficacy population, mean age was 78.9 years in each group. Atrial fibrillation accounted for most anticoagulation indications, and apixaban was the most frequent anticoagulant. Median baseline haematoma volumes were 10.5 and 9.0 mL, median NIHSS scores were nine, and median Glasgow Coma Scale scores were 15 in both groups. Median symptom onset to imaging intervals were 2.3 and 2.4 hours. Thirty-four patients had subdural or subarachnoid haemorrhage enrolled before eligibility narrowed. These features describe a predominantly elderly population with relatively small haemorrhages and preserved consciousness.¹

Among usual care patients in the efficacy population, 195 of 228 (85.5%) received prothrombin complex concentrate, at a median dose of 3000 IU. Where concentrate type was recorded, 92.4% received a four-factor preparation. Most andexanet recipients received the low dose regimen. Treatment crossover was uncommon: four assigned andexanet did not receive it, while three assigned usual care received andexanet.¹

Haemostatic efficacy occurred in 150 of 224 patients (67.0%) assigned andexanet and 121 of 228 (53.1%) assigned usual care. The adjusted difference was 13.4 percentage points (95% CI 4.6–22.2; $P=0.003$). The main contribution came from the imaging component: haematoma enlargement remained within 35% in 76.7% versus 64.6% of evaluable patients. Corresponding proportions without a seven-point NIHSS deterioration were 87.9% and 83.0%; avoidance of rescue treatment occurred in 97.3% and 93.4%.¹

Median anti-factor Xa activity fell by 94.5% with andexanet and 26.9% with usual care ($P<0.001$). These percentages describe the magnitude of assay reduction, not the proportion of patients achieving a clinical response. An analysis including all 530 patients also favoured andexanet for haemostatic efficacy: 63.9% versus 52.4%, an adjusted difference of 11.0 percentage points (95% CI 2.8–19.3).^{1,4}

In the 530-patient population analysed by allocation, thrombotic events occurred in 27 of 263 patients (10.3%) versus 15 of 267 (5.6%): a difference of 4.6 percentage points (95% CI 0.1–9.2; $P=0.048$). Ischaemic stroke occurred in 17 patients (6.5%) versus four (1.5%). Thirty-day deaths numbered 73 (27.8%) and 68 (25.5%), with an adjusted difference of 2.5 percentage points (95% CI –5.0 to 10.0; $P=0.51$). A modified Rankin score of 0–3 was recorded in 69 of 246 assessed andexanet patients (28.0%) and 79 of 255 usual care patients (31.0%). Functional recovery and survival were not demonstrably improved.¹

Critique

Randomisation was a substantial advance over the evidence previously supporting andexanet. It allowed comparison against patients treated at the same sites during the same period, avoiding the major selection and temporal biases that complicate comparisons with historical cohorts. International recruitment strengthened relevance across different systems. Blinded central imaging and event adjudication protected important outcomes from knowledge of treatment, and the coherent separation in anti-factor Xa activity and haematoma expansion makes a real pharmacological and haemostatic effect convincing.¹

The primary endpoint nevertheless asks a narrower question than whether patients recover better. It combined an imaging threshold with avoidance of substantial neurological deterioration and rescue treatment. All components had to be satisfied, but most patients met the clinical and rescue components regardless of allocation. The observed treatment difference was therefore principally an effect on haematoma expansion. A favourable primary result cannot be relabelled as improved neurological outcome simply because an NIHSS condition appeared within the endpoint.^{1,5}

A relative volume threshold also has variable clinical meaning. A similar percentage increase represents very different absolute amounts of blood in a small and a large haematoma. Location and the patient's remaining neurological reserve further influence the consequence of expansion. The supplementary analysis of enlargement by at least 12.5 mL, including early death without repeat imaging, addressed a more substantial absolute change. In the extended population it favoured andexanet, but was outside the main multiplicity-controlled comparison. It supports the biological interpretation without independently establishing a patient centred benefit.⁴

The thrombotic finding belongs alongside efficacy in the principal interpretation. Its confidence interval was imprecise, and the P value for all thrombotic events was close to the conventional threshold. However, the excess was not limited to an asymptomatic laboratory abnormality: ischaemic strokes accounted for a prominent difference. Patients were anticoagulated because they already had thrombotic risk, but randomisation helps distinguish that shared background risk from differences following treatment allocation. It would be inappropriate to dismiss the stroke excess solely because thrombosis can also occur after anticoagulant withdrawal without andexanet.¹

Conversely, it would be inappropriate to subtract the percentage with thrombosis from the percentage with haemostatic success and call the remainder a net benefit. The outcomes differ in severity, timing and definition, may occur in the same patient, and were primarily reported in different analysis populations. Even within a common population, preventing a threshold amount of radiographic expansion and causing a disabling stroke are not interchangeable events. A clinical outcome integrating survival and disability is needed to establish the balance patients experience.

The 30-day functional findings do not supply that definitive answer. Recovery after intracerebral haemorrhage continues beyond a month, functional data were incomplete, and the trial was not powered for disability. These limitations prevent a confident declaration that andexanet has no functional benefit. They equally prevent an assumption that a later benefit must exist. The point estimates did not provide reassurance of better recovery, and the mortality confidence interval remained compatible with clinically important differences in either direction. Longer follow-up and adequate power were needed.^{1,6}

Early stopping further narrows certainty. The interim primary result crossed the reported boundary and was supported by the extended analysis, whose somewhat smaller effect estimate still favoured andexanet. That consistency reduces concern that the finding depended entirely on the first data cut. Stopping nevertheless curtailed the intended sample and accumulated fewer safety and functional observations. The manuscript

reports that the interim boundary was revised before knowledge of the treatment effect. That explanation is relevant, while the reduced information on clinical outcomes remains a consequence of stopping.^{1,4}

The comparator provoked substantive published debate. Some correspondents argued that usual care was weakened by non-use or inadequate dosing of concentrate. The investigators replied that the median weight-adjusted dose was 41 IU/kg, within the commonly used range of 25–50 IU/kg, and that the haemostatic difference persisted when andexanet was compared with usual care patients receiving concentrate. This response matters: the published data do not support the claim that andexanet's entire advantage arose from comparison with patients receiving no concentrate.^{6,6}

That subgroup comparison still cannot replace a trial against a uniform concentrate regimen. Actual concentrate administration within usual care was selected by clinicians, not separately randomised. Dose, preparation and the decision to treat could reflect prognostic factors. ANNEXA-I estimates the effect of offering andexanet against the mixture of usual care used in its participating centres. It does not establish superiority over every possible concentrate dose, nor justify treating a hypothetical better-performing control group as if it had been observed.^{1,6}

Open treatment allocation leaves some scope for differences in co-interventions and decisions about rescue treatment or care limitation. Blinded adjudication reduces measurement bias but cannot remove changes in treatment that occur before an outcome is measured. More rescue treatment in usual care could partly reflect clinician expectations, although its low frequency and the separate imaging result make this an incomplete explanation of the primary finding. Consent-related deletion of 20 randomised patients also means the retained population was not a complete record of everyone randomised; the report describes the reason, but cannot provide their outcomes.¹

Generalisation requires comparable patients and timing. The typical participant was alert, elderly and treated early, with a haematoma far below the maximum permitted volume. Immediate surgical candidates, profound coma and recent thrombosis were excluded. Although multiple countries participated, this does not demonstrate the same benefit–harm balance in these excluded groups. Eligibility also relied substantially on dose timing rather than a universal minimum measured drug concentration. Whether a rapid assay could improve selection by identifying little residual anticoagulant effect remained unresolved.^{1,4,6}

Commercial involvement was extensive and transparent: company personnel contributed to design, investigators had confidentiality agreements, and a sponsor-funded writer assisted with the Methods. Academic coordination, statistical analysis and independent monitoring were important safeguards. These relationships do not negate the randomised results, but reinforce the need to judge the treatment using the full outcome set. The strongest defensible conclusion is specific: andexanet improved early haemostatic control, with an excess of thrombosis and an unresolved effect on outcomes that patients ultimately value.¹

Body of Evidence

The two 2015 volunteer trials, ANNEXA-A and ANNEXA-R, established the pharmacological foundation. A total of 145 healthy older adults receiving apixaban or rivaroxaban were randomised to andexanet or placebo. Anti-factor Xa activity declined rapidly, and bolus-plus-infusion regimens sustained reversal during administration. These trials demonstrated target engagement under controlled conditions. They could not establish whether reversing a drug in a patient with active intracerebral bleeding prevented disability, nor could their small, healthy population reliably quantify uncommon thrombotic complications in patients with atrial fibrillation and acute brain injury.²

ANNEXA-4 then examined patients with acute major bleeding. Its 2019 report included 352 patients, predominantly with intracranial or gastrointestinal bleeding. Excellent or good haemostasis occurred in 204 of 249 evaluable patients (82%), while 34 patients (10%) had a thrombotic event and 49 (14%) died within 30 days. Efficacy assessment required confirmed major bleeding and sufficient baseline anti-factor Xa activity. The study demonstrated feasibility and described outcomes after treatment, but its single-group design could not determine how many haemostatic successes would have occurred under alternative care or how much thrombosis was attributable to andexanet.³

ANNEXA-I therefore changes the certainty of the evidence more than a simple comparison of success percentages suggests. Its lower absolute haemostatic success rate should not be read as a deterioration in drug effectiveness relative to ANNEXA-4: populations, endpoint requirements and analysis rules differed. What matters is that the concurrent randomised control group makes the between-group comparison interpretable. It also places the previously observed approximately 10% thrombotic event rate beside a lower rate under usual care, making safety part of the causal question rather than an unexplained feature of an uncontrolled cohort.^{1,3}

Evidence for concentrate in vitamin K antagonist related haemorrhage addresses a related but distinct mechanism. INCH randomised 54 patients to four-factor concentrate or fresh frozen plasma; 50 received study treatment. Its primary endpoint, an INR no greater than 1.2 within three hours, was achieved in 18 of 27 concentrate recipients (67%) and two of 23 plasma recipients (9%). The study stopped early after a safety review. It favoured concentrate for rapid INR correction, but did not establish a definitive clinical outcome advantage. Restoring deficient coagulation factors during vitamin K antagonist treatment is not the same intervention as countering an active factor Xa inhibitor. INCH consequently informs the rationale for prompt reversal without proving that a particular concentrate strategy would outperform andexanet in ANNEXA-I's population.⁷

The FAST trial of recombinant activated factor VII provides an important warning about relying on haematoma growth. In 841 patients with acute intracerebral haemorrhage, the higher dose reduced expansion compared with placebo but did not improve the primary 90-day clinical outcome. Arterial thromboembolic events were more frequent with the higher dose. FAST tested a different haemostatic agent and cannot determine andexanet's treatment effect. It does, however, demonstrate that a favourable effect on bleeding volume can coexist with failure to improve disability or survival. ANNEXA-I's clinical uncertainty is therefore substantive, rather than an objection that can be resolved by invoking biological plausibility alone.⁸

The 2024 editorial and correspondence concentrated on precisely this gap: demonstrable haemostatic efficacy, uncertain functional benefit and a consequential thrombotic signal. The investigators reasonably defended the randomised efficacy result and the relevance of usual care. Those arguments strengthen confidence in what ANNEXA-I actually tested. They do not transform the 12-hour endpoint into a longer-term benefit, or eliminate the need to characterise which patients gain more from limiting haemorrhage than they lose through thrombotic complications.^{5,6}

At the close of 2024, the trial supported a more informed, selective discussion of andexanet rather than an unqualified claim of superior patient outcomes. For a patient resembling the enrolled population, clinicians could describe a higher probability of early haemostatic control and a higher observed risk of thrombosis. The available evidence could not specify a validated subgroup in whom net benefit was assured. Future trials needed a clearly defined active comparator, longer functional follow-up and sufficient events to examine the clinical trade-off directly. Rapid reversal had become better evidenced; improved recovery remained an unanswered question.¹

Does andexanet improve outcomes in factor Xa inhibitor associated intracerebral haemorrhage?

Andexanet improves early haemostatic control but increases thrombotic events, including ischaemic stroke. ANNEXA-I did not establish a survival or functional benefit, leaving the overall clinical benefit uncertain.

References

1. Connolly SJ, Sharma M, Cohen AT, Demchuk AM, Członkowska A, Lindgren AG, et al. Andexanet for Factor Xa Inhibitor–Associated Acute Intracerebral Hemorrhage. *N Engl J Med* 2024;390(19):1745–1755.
2. Siegal DM, Curnutte JT, Connolly SJ, Lu G, Conley PB, Wiens BL, et al. Andexanet Alfa for the Reversal of Factor Xa Inhibitor Activity. *N Engl J Med* 2015;373(25):2413–2424.
3. Connolly SJ, Crowther M, Eikelboom JW, Gibson CM, Curnutte JT, Lawrence JH, et al. Full Study Report of Andexanet Alfa for Bleeding Associated with Factor Xa Inhibitors. *N Engl J Med* 2019;380(14):1326–1335.
4. ANNEXA-I Investigators. Supplementary Appendix to: Andexanet for Factor Xa Inhibitor–Associated Acute Intracerebral Hemorrhage. *N Engl J Med*. doi:10.1056/NEJMoa2313040.
5. Smith WS, Hemphill JC. Reversing Oral Anticoagulation in Intracerebral Hemorrhage. *N Engl J Med* 2024;390(19):1815–1816.
6. Andexanet for Factor Xa Inhibitor–Associated Acute Intracerebral Hemorrhage [correspondence and authors’ reply]. *N Engl J Med* 2024;391(8):772–774. doi:10.1056/NEJMc2407378.
7. Steiner T, Poli S, Griebel M, Hüsing J, Hajda J, Freiburger A, et al. Fresh frozen plasma versus prothrombin complex concentrate in patients with intracranial haemorrhage related to vitamin K antagonists (INCH): a randomised trial. *Lancet Neurol* 2016;15(6):566–573.
8. Mayer SA, Brun NC, Begtrup K, Broderick J, Davis S, Diringer MN, et al. Efficacy and Safety of Recombinant Activated Factor VII for Acute Intracerebral Hemorrhage. *N Engl J Med* 2008;358(20):2127–2137.

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.

Introduction

Removing an intracerebral haematoma seems an intuitive response to a destructive collection of blood. Evacuation can reduce mass effect and might limit secondary injury from blood products. The operation also requires access through brain tissue, however, and can cause additional injury or recurrent bleeding. The clinical question is whether the benefit of removing blood exceeds the harm of reaching and evacuating it. Earlier randomised trials did not establish an overall functional advantage from conventional surgery, leaving uncertainty about which patients, techniques and treatment windows could produce a better balance.¹

Haemorrhage location is central to that balance. A lobar collection may be more accessible than a deep haemorrhage embedded within structures and white matter pathways that are essential to function. Timing may also matter: an operation performed after secondary injury develops might have different consequences from early evacuation. STICH and STICH II suggested that selection by anatomical location deserved further investigation, while MISTIE III tested a different strategy of catheter evacuation followed by thrombolysis. These approaches could not be treated as interchangeable tests of a single operation.^{2,3,4}

ENRICH investigated early trans-sulcal parafascicular evacuation through a minimally invasive access port. The intervention combined image-guided access, immediate removal of a substantial proportion of the haematoma and standardised neurocritical care. It also used an adaptive design that could stop recruitment within an anatomical subgroup if accumulating results suggested insufficient benefit.¹

The trial's positive primary result was a major development because it concerned functional outcome at six months. Its interpretation nevertheless depends on the selected population and the adaptation that occurred during recruitment. The evidence principally supports patients with lobar haemorrhage who resemble those enrolled and can receive the studied operation promptly. It does not establish a general benefit from surgery for all intracerebral haemorrhages or prove superiority of this device over another surgical technique.¹

Synopsis

ENRICH was a prospective, multicentre, randomised trial conducted at 37 centres in the United States. Between December 2016 and August 2022, 11,603 patients were assessed and 300 were randomised, with 150 assigned to minimally invasive surgery plus medical management and 150 to medical management alone. Recruitment was interrupted during the COVID-19 pandemic. The scientific leadership team at Emory University oversaw the trial; statisticians from Berry Consultants designed its adaptive methodology and performed the analyses.¹

Nico, the manufacturer of the BrainPath access port and Myriad evacuation device, funded the study. The report states that the sponsor did not design the scientific trial, collect or analyse the data, or decide whether to submit the manuscript. Investigators had independent access to the database and no confidentiality agreements with the sponsor. Participating sites purchased the devices through their usual clinical arrangements. Surgeons had previous experience with the procedure and completed prerequisite manufacturer-organised training.^{1,5}

Eligible patients were 18–80 years old, with a spontaneous supratentorial haemorrhage estimated locally at 30–80 mL, a Glasgow Coma Scale score of 5–14 and little or no premorbid disability, defined as a modified Rankin score of 0 or 1. The principal report required an NIH Stroke Scale score above five. Surgery had to be feasible within 24 hours of the time the patient was last known well. Lobar and anterior basal ganglia haemorrhages were initially eligible; primary thalamic and infratentorial haemorrhages were excluded.¹

Other exclusions included extensive intraventricular blood occupying more than half of either lateral ventricle, a secondary cause of haemorrhage, uncorrectable coagulopathy and a need to resume long-term anticoagulation within five days. Patients already designated for comfort care, or without a reasonable expectation of recovery, were also excluded. The study therefore addressed a selected population with substantial haemorrhage but a prospect of recovery, rather than every patient considered for lifesaving neurosurgery.^{1,5}

Central online randomisation was stratified by haemorrhage location and Glasgow Coma Scale score below nine or at least nine. Surgery involved general anaesthesia, a small craniotomy and image-guided access through a sulcus. The BrainPath port provided a working channel, and the haematoma was removed with suction and the Myriad device. The surgeon achieved haemostasis and replaced the bone flap. This was an evacuation procedure, distinct from decompressive craniectomy alone.¹

Both groups received medical management guided by a common clinical standardisation document. Operations were recorded; the first two at each site underwent central review, with additional cases reviewed when needed. Crossover to the study operation was prohibited, but lifesaving conventional craniotomy or decompressive hemicraniectomy remained available. The comparison was consequently a policy of early minimally invasive evacuation against initial medical management with rescue surgery when clinically necessary.¹

Patients and clinical teams were unblinded. Local staff obtained structured, recorded modified Rankin interviews at 30, 90 and 180 days, sometimes using telephone interviews or caregiver information. A third party removed references to treatment allocation, and an independent neuropsychologist adjudicated the redacted recordings. Imaging underwent central review, although not all imaging readers were blinded. The independent physician reviewing serious adverse events and deaths knew allocation.¹

The primary efficacy endpoint was the mean utility-weighted modified Rankin score at 180 days. Conventional scores from zero to six were assigned values of 1.0, 0.91, 0.76, 0.65, 0.33, zero and zero respectively; higher utility scores represented better outcomes. The primary safety endpoints were death within 30 days and change in haematoma volume. Secondary measures included the ordinal modified Rankin scale, functional outcomes dichotomised at scores 0–3 versus 4–6, lengths of stay and postoperative rebleeding with neurological deterioration, defined by an NIH Stroke Scale increase of at least four or a GCS decrease of at least two points.¹

The Bayesian design permitted 150–300 participants and repeated interim analyses. Superiority at the final primary analysis required a posterior probability of at least 0.975 that surgery improved mean utility. Recruitment within a location could stop for futility if the estimated probability of a clinically meaningful benefit greater than 0.075 was sufficiently low. After 175 patients had entered, this rule stopped anterior basal ganglia recruitment. Lobar recruitment continued to the maximum sample of 300, ultimately comprising 208 lobar and 92 anterior basal ganglia haemorrhages.^{1,5}

All randomised patients were included in the primary analysis, with Bayesian imputation for missing 180-day scores using earlier outcomes. The primary superiority decision used a pooled model, while a complementary hierarchical model estimated effects by location. Simulations described in the report suggested approximately 90% or greater power for a utility difference of 0.15 across both locations and an approximate one-sided type I error below 0.025. Secondary and exploratory analyses were not adjusted for multiplicity.¹

Median age was 64 years with surgery and 62 with control. Median haematoma volumes were 54 and 55 mL, and median NIH Stroke Scale scores were 16 and 18. Approximately four fifths had a Glasgow Coma Scale score of at least nine. Surgery began a median 16.75 hours after the patient was last known well. Observed 180-day Rankin scores were available for 147 surgical patients and 139 controls, representing 95.3% of the total cohort.¹

The primary mean utility score was 0.458 with surgery and 0.374 with control: a difference of 0.084 (95% Bayesian credible interval 0.005–0.163), with posterior probability of superiority 0.981. The location-specific estimated difference was 0.127 (95% credible interval 0.035–0.219) for lobar haemorrhage and –0.013 (–0.147 to 0.116) for anterior basal ganglia haemorrhage. Among assessed patients, a conventional Rankin score of 0–3 occurred in 74 of 147 (50.3%) versus 57 of 139 (41.0%).¹

Mean haematoma reduction after surgery was 73.2%, and 109 patients had a reported residual volume no greater than 15 mL. Death within 30 days occurred in 14 of 150 surgical patients (9.3%) and 27 of 150 controls (18.0%), an estimated difference of –8.7 percentage points (95% credible interval –16.4 to –1.0). By 180 days, deaths numbered 30 and 35. Five surgical patients (3.3%) had postoperative rebleeding with neurological deterioration. Serious adverse events affected 95 (63.3%) and 118 (78.7%) patients respectively; decompressive hemicraniectomy was performed in five and 30.¹

Critique

ENRICH's most important strength was its direct assessment of functional recovery. A substantial reduction in haematoma volume was accompanied by a positive six-month outcome, rather than being used as its substitute. Central randomisation, balanced allocation, standardised clinical care and adjudicated functional assessments strengthen the inference that the treatment strategy contributed to the difference. The large number of participating surgeons also shows that the intervention was not confined to the experience of one exceptional operator.¹

The operation was nevertheless delivered within a trained and monitored network. Previous procedural experience, a common manual and review of recorded operations were integral to the tested intervention. They support consistency while limiting the assumption that buying the equipment alone reproduces the result. The relevant clinical service needs timely imaging, suitable patient selection, theatre access, experienced surgeons and postoperative neurocritical care. ENRICH supports an organised pathway for a defined population, not a device detached from the expertise and care surrounding it.¹

The adaptive design addressed a clinically plausible source of heterogeneity. Stopping anterior basal ganglia recruitment was a reported design feature, rather than simply removing an inconvenient subgroup from the final publication. All 92 previously enrolled patients from that location remained in the primary analysis. The study then completed its maximum overall sample. Describing ENRICH as a trial stopped early for benefit would therefore be incorrect; recruitment was enriched for lobar haemorrhage after a location-specific futility decision.^{1,5}

Futility does not establish that the operation can never help any patient with anterior basal ganglia haemorrhage. It indicates that the accumulating evidence did not meet the trial's threshold for continuing that subgroup. Its final estimate was close to no effect, with a wide interval allowing both harm and benefit. The positive pooled result is chiefly relevant to lobar haemorrhage, which comprised approximately 70% of the final cohort. The pooled estimate should not become a universal average applied to deep haemorrhages that stopped entering the study.¹

The Bayesian result also needs precise language. A posterior probability of 0.981 is a probability of a positive mean effect under the specified model and prior assumptions. It is not a 98.1% probability that any individual patient benefits, nor evidence that almost all patients recover. The estimated utility gain of 0.084 is not an 8.4 percentage point increase in independence. Its interval includes very small gains and larger benefits, and the threshold for declaring superiority concerned an effect above zero rather than one necessarily exceeding the 0.075 value used in the futility rule.¹

Utility weighting recognises that steps on an ordinal disability scale may differ in importance. It also embeds preferences within the endpoint. In ENRICH, severe disability at Rankin score five and death both received zero, while the change from score four to three received appreciably more weight than the change from one to zero. These choices influence what contributes to the mean effect. The weights had precedents in stroke research but had not been specifically validated for this intracerebral haemorrhage application. Individual patients may value survival with major dependency differently from the scale's assigned values.¹

The ordinary Rankin results help make the benefit clinically intelligible. Both the ordinal analysis and the observed proportion reaching scores 0–3 favoured surgery. However, the dichotomous difference remained uncertain, and these additional analyses were outside a multiplicity-adjusted confirmatory framework. Agreement across related outcomes is supportive; it does not convert every favourable secondary estimate into a separate proven benefit. The six-month primary result should remain the basis of the efficacy claim.¹

Missing functional outcomes were relatively uncommon but uneven: three in the surgical arm and 11 in control. Imputation retained all 300 participants and used earlier functional information, an appropriate way to use observed follow-up rather than simply discarding incomplete cases. Its validity still depends on assumptions about why data were missing. The supplement acknowledges those assumptions, particularly for the six patients missing both three- and six-month outcomes. A small missing fraction does not make the problem disappear, particularly when the primary credible interval lies close to zero at its lower limit.^{1,5}

Outcome masking was stronger than an entirely unblinded bedside assessment, but imperfect. Local interviewers knew allocation before recordings were redacted, so their questioning and information gathering could already have been influenced. Blinded central adjudication cannot recreate information that was not elicited. Likewise, unblinded clinical teams could alter rehabilitation, escalation and rescue decisions. Such possibilities warrant caution without assuming that they explain the observed functional difference. The common care framework and independent adjudication were meaningful safeguards.¹

The early mortality difference is encouraging, but its attenuation by six months matters. ENRICH does not establish that the entire 30-day survival advantage persists indefinitely. Postoperative rebleeding caused deterioration in five patients, and cardiac arrests were more frequent in the surgical group, mostly after discharge. Those small event counts cannot reliably identify a mechanism or prove causation. They do show why a positive efficacy result should be presented alongside procedure-related complications and the full follow-up, rather than accompanied by an unqualified description of surgery as safe.^{1,5}

Length of stay and ventilator duration require particular restraint. The supplementary analyses excluded patients who died before discharge, conditioning these comparisons on an event affected by treatment. Because early deaths differed between groups, the surviving patients need not remain comparable despite randomisation. Shorter stays among survivors may be useful descriptive information, but cannot by themselves establish reduced overall resource use or a hospital-wide economic benefit. A formal economic analysis would also need procedural costs and outcomes after discharge.⁵

Generalisability is narrow in other important respects. Only 300 of more than 11,000 screened patients entered; many were excluded for volume, location or severity, and some because basal ganglia recruitment had closed. Participants were previously independent, no older than 80, and had an operation feasible within 24 hours. Results cannot simply be transferred to extensive intraventricular haemorrhage, thalamic or

posterior fossa bleeding, very small or very large haematomas, or substantial premorbid disability. Sponsor funding warrants disclosure, but the stated academic control and independent data access are strengths. The remaining implementation question is whether similarly selected patients can obtain comparable outcomes beyond this experienced US network.^{1,5}

Body of Evidence

STICH tested a broad policy of early evacuation against initial conservative treatment with later surgery if needed. It randomised 1033 patients across 83 centres. Favourable six-month outcomes occurred in 26% of assessed surgical patients and 24% of controls, without a significant overall benefit. The operation was scheduled within 24 hours of randomisation, a different clock from ENRICH's requirement to operate within 24 hours of last known wellness. STICH established the need for better selection and technique rather than confirming that every form of evacuation was ineffective.²

STICH II concentrated on conscious patients with superficial lobar haemorrhages of 10–100 mL and no intraventricular haemorrhage. Among 601 randomised patients, unfavourable six-month outcomes occurred in 174 of 297 assessed surgical patients (59%) and 178 of 286 controls (62%). The odds ratio was 0.86 (95% CI 0.62–1.20), leaving the primary comparison inconclusive. Its anatomical focus helped motivate ENRICH, but its population, conventional surgical approach and prognosis-based Glasgow Outcome Scale endpoint differed. The two trials cannot establish which individual change in ENRICH produced the later positive result.³

MISTIE III provided a large test of catheter evacuation followed by repeated small doses of alteplase. It randomised 506 patients with supratentorial haemorrhage of at least 30 mL, with 499 entering its modified intention-to-treat efficacy analysis. The adjusted estimated proportion reaching a Rankin score of 0–3 at one year was 45% with intervention and 41% with medical care, a risk difference of four percentage points (95% CI –4 to 12; $P=0.33$). This did not demonstrate the primary functional benefit. ENRICH differs through direct removal during an early operation, but it did not randomise patients between that approach and MISTIE. Cross-trial comparison therefore generates explanations about timing, location and evacuation efficiency without establishing their causal contributions.⁴

The smaller ICES study explored CT-guided endoscopic evacuation. Its randomised cohort contained 20 patients, and analyses also incorporated medical controls from the contemporaneous MISTIE programme. It showed that substantial early evacuation could be achieved, with encouraging but imprecise functional findings. That design was useful

for feasibility and planning, but was too small to settle patient benefit. ENRICH advanced beyond this exploratory evidence by evaluating a reproducible surgical strategy in a larger randomised cohort with a functional primary endpoint.⁶

Together, these studies argue against treating intracerebral haemorrhage surgery as a single intervention. Lobar accessibility, deep tissue injury, the amount of residual blood, the timing of intervention and the possibility of rescue surgery all change the clinical comparison. ENRICH's association between greater evacuation and better outcome is biologically plausible, but patients were not randomised to different residual volumes. Easier anatomy or less severe tissue damage could influence both successful removal and recovery. The association should guide further investigation and procedural quality monitoring, rather than becoming proof that reaching a particular volume guarantees benefit.^{1,7}

The accompanying editorial regarded ENRICH as a reason for renewed optimism in lobar haemorrhage and emphasised effective evacuation within locally organised surgical services. That enthusiasm is compatible with a cautious interpretation: the trial offers evidence for an actual treatment option in a field with several neutral trials, while leaving its scope defined by the studied anatomy and setting. It supports early specialist discussion for comparable patients, not automatic surgery solely because a haemorrhage has occurred.^{1,7}

By the end of 2024, ENRICH had established a credible functional benefit from an early minimally invasive evacuation strategy in its selected cohort, principally attributable to lobar haemorrhage. It left unresolved the best approach for deep haemorrhages, comparative performance against other operations, the optimal treatment window within the first day and the effect of broader implementation. Those uncertainties should shape further trials and service evaluation. They should neither erase the positive result nor expand it beyond the patients and treatment pathway that generated it.¹

Does early minimally invasive evacuation improve recovery after intracerebral haemorrhage?

ENRICH supports improved six-month functional outcome with the studied operation for selected lobar haemorrhages of 30–80 mL treated within 24 hours. Benefit was not established for anterior basal ganglia haemorrhage or for other surgical techniques.

References

1. 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.
2. 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.
3. 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.
4. 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.
5. ENRICH Trial Investigators. Supplementary Appendix to: Trial of Early Minimally Invasive Removal of Intracerebral Hemorrhage. *N Engl J Med*. doi:10.1056/NEJMoa2308440.
6. Vespa P, Hanley D, Betz J, Hoffer A, Engh J, Carter R, et al. ICES (Intraoperative Stereotactic Computed Tomography-Guided Endoscopic Surgery) for Brain Hemorrhage: A Multicenter Randomized Controlled Trial. *Stroke* 2016;47(11):2749–2755.
7. Mendelow AD. New Hope for Adults with Lobar Intracerebral Hemorrhage. *N Engl J Med* 2024;390(14):1328–1329.

Possible Reporting Inconsistencies. Table 3 defines neurological deterioration after rebleeding using a GCS decrease “of up to 2 points”, whereas the Methods and supplementary Table S9 specify a decrease of at least two points. The wording in Table 3 appears reversed; the latter definition is used here. This discrepancy does not establish that the reported event count is incorrect.^{1,5}

SWITCH

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.

Introduction

A large deep intracerebral haemorrhage creates two related problems: injury within structures essential for neurological function, and mass effect within a rigid skull. Removing the clot might reduce pressure and exposure to blood products, but reaching a deep collection can damage additional tissue. Decompressive craniectomy offers a different approach. Removing a large bone flap creates space for the swollen brain without requiring the surgeon to enter and evacuate the haematoma. Whether that physiological rationale produces an outcome patients would value remained uncertain.¹

Evidence from malignant middle cerebral artery infarction supported decompression as a means of preventing death from swelling. However, an infarct and a deep haemorrhage are not interchangeable lesions. The tissue initially damaged, subsequent evolution and risks of recurrent bleeding differ. Survival after either event can involve considerable dependency, making the distribution of functional outcomes at least as important as the mortality rate.^{2,3}

Trials of haematoma evacuation had also left a specific gap. Conventional surgery in STICH and STICH II did not establish an overall functional benefit, while MISTIE III's catheter-based strategy did not meet its primary functional endpoint. By 2024, ENRICH had reported benefit from early minimally invasive evacuation principally in lobar haemorrhage, but its anterior basal ganglia cohort had stopped recruiting for futility. None directly answered whether decompression without evacuation could help selected patients with severe deep bleeding.^{4,5,6,7}

SWITCH therefore tested decompressive craniectomy alongside best medical treatment against best medical treatment alone. It deliberately focused on a severely affected but previously independent population and used an endpoint separating death or the most severe dependency from survival with less dependency. The hypothesis was that relieving pressure could prevent devastating outcomes even when removal of the underlying deep injury was impossible.¹

Synopsis

SWITCH was an academic, multicentre, randomised, open-label trial with blinded outcome assessment, conducted across 42 stroke centres in nine European countries. Recruitment began in October 2014, and the last participant entered in April 2023. The investigators randomised 201 patients; four subsequently declined delayed consent and their records were removed. The full analysis set therefore contained 197 participants, with 96 allocated to decompressive craniectomy plus best medical treatment and 101 to best medical treatment alone. An additional observational cohort was separate from the randomised comparison.¹

The intended sample was 300. Interim reviews after 100 and 150 participants supported continuation, but the trial formally stopped in April 2023 because further funding was unavailable. Support came from the Swiss National Science Foundation, Swiss Heart Foundation, Inselspital Stiftung and an unrestricted Boehringer Ingelheim grant. The report states that funders had no role in design, data collection, analysis, interpretation or manuscript preparation. The principal investigators remained masked to group-specific clinical outcomes until stopping.¹

Eligibility required age 18–75 years, a spontaneous deep supratentorial haemorrhage involving the basal ganglia or thalamus, a stable haematoma of 30–100 mL, a Glasgow Coma Scale score of 8–13 and an NIH Stroke Scale score of 10–30. Patients had a premorbid modified Rankin score of zero or one. Extension into lobar tissue, ventricles or the subarachnoid space was permitted, but exclusively lobar and infratentorial haemorrhages were excluded. Randomisation had to occur within 66 hours of onset, with surgery feasible within the next six hours.^{1,8}

Secondary causes of haemorrhage, substantial pre-existing disability, dementia and an expectation of death within 24 hours were among the exclusions. Coagulation had to be adequately corrected, including an international normalised ratio below 1.5 and platelet count above $100 \times 10^9/L$. The trial thus selected patients with a severe lesion and impaired consciousness, while excluding both milder presentations and the deepest coma. Screening logs were not maintained, so the proportion of otherwise eligible patients entering the trial cannot be established.^{1,8}

Central web-based allocation used probabilistic minimisation for age, baseline stroke severity, expected time to decompression and centre. The operation required a craniectomy at least 12 cm in diameter, without haematoma evacuation. Reimplantation of the bone flap was planned after one to five months. Both groups received guideline-based medical treatment. Patients and treating teams knew assignment, while certified assessors used structured telephone interviews for the six- and twelve-month

Rankin outcomes and an independent core laboratory assessed imaging while masked to allocation.¹

The primary outcome was a modified Rankin score of five or six at 180 days: the most severe disability or death. Secondary outcomes included the full ordinal scale, alternative dichotomisations, mortality, neurological scores, quality of life, permanent residential care, length of hospital stay and imaging measures. Serious adverse events were assessed separately, with safety analyses grouping patients according to treatment received. This produced different safety denominators from the randomised efficacy groups.^{1,8}

The sample calculation assumed a 53% primary-event rate with medical treatment and a one-third relative reduction with surgery, giving more than 85% power with 300 participants. The principal comparison was stratified for minimisation factors except centre, because many centres enrolled few patients. As fewer than 5% of primary outcomes were missing, the reported primary analysis used complete cases. Sensitivity analyses included multiple imputation and a per-protocol population. Secondary analyses were exploratory and had no adjustment for multiple comparisons.^{1,8}

Median age was 61 years, 32% were women, median baseline Glasgow Coma Scale score was ten, and median haematoma volume was 57 mL. Randomisation occurred a median 24 hours after onset; decompression occurred at 26 hours, with a median diameter of 13 cm. By day seven, eight patients assigned to medical treatment had crossed to decompression and one assigned to surgery had not received it. Haematoma evacuation was performed in nine surgical-arm patients and eight controls during that period, despite not forming part of the assigned decompression procedure.¹

At 180 days, primary outcomes were available for 95 patients in each group. A Rankin score of five or six occurred in 42 of 95 (44%) assigned to decompression and 55 of 95 (58%) assigned to medical treatment: adjusted risk ratio 0.77 (95% confidence interval 0.59–1.01; $P=0.057$). The appendix gives the adjusted risk difference more precisely as –13.1 percentage points (95% confidence interval –26.2 to 0.5). The per-protocol estimate was similar: 36 of 77 versus 44 of 73, adjusted risk ratio 0.76 (0.58–1.00). With multiple imputation, the adjusted risk ratio was 0.76 (0.59–0.99; $P=0.042$).^{1,8}

By 180 days, 16 surgical-arm patients and 27 controls had died; the adjusted mortality risk ratio was 0.61 (0.36–1.01; $P=0.065$). The imputed ordinal analysis favoured surgery, with a common odds ratio for a worse Rankin outcome of 0.57 (0.34–0.97). However, the estimated proportion with scores four to six was 86% in both groups. At one year, the

adjusted risk ratio for scores five or six remained compatible with benefit or no benefit, at 0.81 (0.60–1.08).¹

Serious adverse events affected 42 of 103 patients receiving decompression (41%) and 41 of 94 receiving medical treatment (44%), a risk difference of –3 percentage points (95% confidence interval –16 to 11). Permanent cerebrospinal fluid shunting at six months was estimated in 19% versus 8% of survivors, with an imprecise adjusted risk ratio of 2.30 (0.85–6.26). Quality-of-life and residential-care outcomes did not demonstrate a broad advantage. Of 62 surviving surgical patients who answered whether they would undergo the operation again, 48 (77%) said yes.^{1,8}

Critique

SWITCH addressed a consequential question that observational surgical series could not reliably settle. Decisions to operate depend on prognosis, anatomy, deterioration and clinicians' expectations, all of which also influence outcome. Randomisation substantially improves on comparing selected operated patients with non-operated controls. Central allocation, blinded functional assessment, independent imaging review and a primary endpoint measured at six months are important strengths. The investigators also preserved the distinction between decompression and clot removal, allowing the result to be interpreted as evidence for a specific treatment strategy.¹

The major limitation is precision. Funding ended recruitment after approximately two thirds of the planned sample. The primary point estimate suggests a clinically important reduction in death or extreme dependency, but its confidence interval includes no effect. Calling the operation ineffective solely because P was 0.057 would discard that uncertainty; calling the result definitively positive would discard it in the opposite direction. The defensible conclusion is suggestive evidence of benefit requiring confirmation. Stopping for lack of funding also differs from stopping because accumulating outcomes established efficacy or futility.¹

The assumed treatment effect was large. Even the planned study was designed around a one-third relative reduction in the primary outcome, rather than a modest effect that patients might still consider worthwhile. The incomplete sample made such smaller benefits especially difficult to distinguish from chance. The long recruitment period illustrates the practical difficulty of conducting a surgical trial in a narrowly selected emergency population. It does not, by itself, invalidate the randomised comparison, but medical care and referral pathways could evolve over almost nine years.¹

Missing primary data were uncommon overall but asymmetric: one surgical patient and six controls lacked the six-month assessment. Multiple imputation produced almost the same effect estimate while moving its interval just below one. These results support consistency in direction, rather than representing two fundamentally different answers. The primary complete-case analysis remains the designated comparison, and the imputation result depends on assumptions about missing outcomes. Likewise, the supplementary confidence distribution's 97.2% confidence in a risk ratio below one is a post-hoc presentation of the same limited data, not an independent confirmatory result or a prespecified Bayesian posterior.^{1,8}

The endpoint carries a particularly important value judgement. Avoiding Rankin scores five and six includes reaching score four, a state of moderately severe disability requiring substantial assistance. At six months, the observed distribution contained 40 patients at score four after assignment to decompression and 26 after medical treatment; the numbers at scores zero to three were 13 and 14. The trial therefore offers little evidence that decompression increases the chance of a broadly independent recovery. Its potential benefit concerns avoiding death or greater dependency while accepting that major disability is common.¹

A shift from score five to four may be highly meaningful to a patient and family, as may survival itself. It should not be dismissed merely because independence is unlikely. Conversely, a clinician cannot assume that every patient would accept the likely surviving health state. The 77% willingness to repeat surgery is useful information from selected survivors who answered the question. It cannot represent the preferences of all randomised patients, those who died or those unable or unwilling to respond. Early discussions need to make the possible outcomes intelligible and incorporate the person's prior values.^{1,9}

Treatment separation was substantial but incomplete. Rescue decompression in controls can dilute the contrast between initial strategies, while additional clot evacuation introduces variation in subsequent care. These are clinically understandable events in severe brain injury. Nevertheless, one cannot simply claim that a larger benefit would have appeared without crossover: those crossing over are selected by clinical events after randomisation. Per-protocol and as-treated comparisons lose some protection from random allocation, even when they point in the same direction. The intention-to-treat estimate remains the most credible guide to offering the initial strategy.^{1,8}

Clinical masking was impossible, and blinded telephone assessment cannot remove every effect of unblinded care. Treatment allocation could influence escalation,

rehabilitation or expectations about prognosis. The small sample also limits the reliability of subgroup findings. Absence of a statistically supported interaction does not establish equal benefit across ages, haematoma volumes, neurological severities or precise deep locations. Exploratory associations with timing and craniectomy size are particularly vulnerable to confounding because neither feature was independently randomised.^{1,8}

The safety findings do not prove equivalence. Their wide intervals allow clinically relevant differences, and grouping patients by received treatment changes the comparison from a fully randomised one. Cerebrospinal fluid shunting was numerically more frequent after surgery, consistent with a possible treatment burden even though the estimate was uncertain. Decompression also commonly entails a later bone-flap procedure. The decision therefore includes operative complications, further procedures and prolonged recovery, as well as the chance of avoiding death or the worst dependency.¹

Secondary outcomes reinforce the need for measured claims. The favourable imputed ordinal result was not part of a multiplicity-adjusted confirmatory framework. Broad quality-of-life and permanent-care benefits were not established, and follow-up information was less complete for several of these measures than for the primary endpoint. Length of stay among survivors conditions the analysis on survival, which treatment might affect; it cannot establish an overall resource saving. Reduced early midline shift supports the proposed mechanical action, but does not prove prevention of permanent tissue injury or explain the clinical effect.^{1,8}

Applicability is confined to patients resembling those enrolled: previously independent adults no older than 75, with stable, large deep haemorrhages and intermediate rather than profound coma, treated within the specified window. The evidence cannot be transferred directly to lobar haemorrhage, posterior fossa bleeding, much smaller collections or severe premorbid disability. Without screening logs, the degree of selection at participating centres remains uncertain. SWITCH can inform an individual multidisciplinary discussion in this difficult population, but does not establish routine decompression for every severe intracerebral haemorrhage.^{1,8}

Body of Evidence

The strongest precedent for decompression came from ischaemic stroke. A pooled analysis of DECIMAL, DESTINY and HAMLET included 93 patients aged 18–60 with malignant middle cerebral artery infarction treated within 48 hours. At one year, survival was 78% with surgery and 29% with medical treatment; a Rankin score of four or better occurred in 75% and 24%. The pooled absolute improvement for the latter outcome was 51 percentage points (95% confidence interval 34–69). This established that decompression could transform outcomes in a selected swelling-related brain injury. It did not establish an equivalent effect in deep intracerebral haemorrhage, where the initial lesion and subsequent risks differ.²

DESTINY II extended the ischaemic-stroke question to 112 patients aged at least 61. Assignment to hemicraniectomy within 48 hours improved estimated six-month survival with Rankin scores zero to four from 18% to 38%, with an odds ratio of 2.91 (95% confidence interval 1.06–7.49). These were estimates corrected for the sequential design, rather than simple raw proportions. Mortality was lower, but no assessed survivor reached scores zero to two and most survivors required substantial assistance. The relevance to SWITCH is therefore both mechanistic and ethical: a successful lifesaving intervention may leave considerable disability, requiring the endpoint to be translated into outcomes that patients can understand.³

The haemorrhage evacuation trials addressed a different mechanism. STICH enrolled 1033 patients in a broad comparison of early surgery with initial conservative treatment and did not demonstrate an overall functional advantage. STICH II narrowed selection to superficial lobar haemorrhage without intraventricular blood, but its 601-patient trial also left the primary comparison inconclusive. Neither trial directly tested the SWITCH operation. Their results discourage an undifferentiated claim about the benefits of “surgery”, while leaving room for a distinct intervention in a different anatomical population.^{4,5}

MISTIE III randomised 506 patients to catheter evacuation with alteplase or medical care. Its modified intention-to-treat analysis did not demonstrate improvement in the primary one-year functional endpoint. The procedure aimed to remove blood gradually, whereas SWITCH predominantly left the clot in place and relieved external constraint. Consequently, MISTIE's neutral primary result cannot be used as a direct refutation of decompression, nor can SWITCH establish that decompression is superior to catheter evacuation.⁶

ENRICH supplied a complementary result in 2024. Its 300-patient trial found a higher mean utility-weighted Rankin score at six months after early minimally invasive

evacuation. The effect was principally attributable to lobar haemorrhage, while anterior basal ganglia recruitment stopped for futility. SWITCH concentrated on deep haemorrhage and used a different operation, treatment window and primary endpoint. These trials support separating anatomical and procedural questions. They do not provide a head-to-head comparison or establish a combined evacuation-and-decompression strategy.^{1,7}

The accompanying SWITCH commentary emphasised the tentative nature of the benefit and the substantial disability among survivors. It also considered the difficulties of translating an uncertain average effect into a surgical decision for an individual patient. A broader Lancet Neurology editorial welcomed renewed progress in haemorrhagic-stroke research while highlighting the need for sustained investment. SWITCH's premature funding-related closure gives that argument practical force: an important question can remain unresolved even after years of multinational recruitment and careful follow-up.^{9,10}

By the end of 2024, the evidence supported a more specific conversation about severe deep haemorrhage. Decompressive craniectomy might reduce death or the most severe dependency in carefully selected patients, but the size and certainty of benefit remained insufficiently established. Future work needs adequate recruitment, more complete assessment of surviving health states and caregiver burden, and clearer identification of patients whose preferences and prognosis justify the intervention. Until then, the trial's contribution is meaningful but limited: it supplies randomised evidence for a difficult decision without making that decision automatic.^{1,9}

Does decompressive craniectomy improve outcome after severe deep intracerebral haemorrhage?

SWITCH suggests a possible reduction in death or extreme dependency in selected patients, but the primary result was inconclusive. Substantial disability remained common, making individual preferences central to any decision to operate.

References

1. 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.

2. Vahedi K, Hofmeijer J, Juettler E, Vicaut E, George B, Algra A, et al. Early decompressive surgery in malignant infarction of the middle cerebral artery: a pooled analysis of three randomised controlled trials. *Lancet Neurol* 2007;6(3):215–222.
3. Jüttler E, Unterberg A, Woitzik J, Bösel J, Amiri H, Sakowitz OW, et al. Hemispherectomy in older patients with extensive middle-cerebral-artery stroke. *N Engl J Med* 2014;370(12):1091–1100.
4. 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.
5. 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.
6. 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.
7. 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.
8. SWITCH Study Investigators. Supplementary appendix to: Decompressive craniectomy plus best medical treatment versus best medical treatment alone for spontaneous severe deep supratentorial intracerebral haemorrhage. *Lancet*. doi:10.1016/S0140-6736(24)00702-5.
9. Klijn CJM, Dammers R, Sprigg N. Decompressive craniectomy for deep intracerebral haemorrhage: a SWITCH towards better outcomes? *Lancet* 2024;403(10442):2351–2353.
10. The Lancet Neurology. Forging ahead in haemorrhagic stroke research. *Lancet Neurol* 2024;23(6):545. doi:10.1016/S1474-4422(24)00181-9.

INTREPID

Greer DM, Helbok R, Badjatia N, Ko SB, Guanci MM, Sheth KN; INTREPID Study Group. Fever Prevention in Patients With Acute Vascular Brain Injury: The INTREPID Randomized Clinical Trial. JAMA 2024;332(18):1525–1534.

Introduction

Fever after a severe stroke is associated with worse neurological outcomes. Increased metabolic demand in injured tissue provides a plausible reason to prevent it, but fever also accompanies infection, inflammation and the severity of the original lesion. An association with poor recovery does not establish that lowering temperature will improve recovery. This distinction matters when treatment requires a cooling device, additional monitoring and medication to control shivering.¹

Preventing fever also differs from inducing hypothermia. A target of 37.0°C aims to maintain normal temperature, while treatment begun only after fever develops accepts some exposure before intervening. Both approaches may involve active temperature management. Their comparison therefore asks whether earlier, more consistent control adds benefit to a strategy that already treats established fever, rather than whether high temperature should simply be ignored.¹

Earlier randomised work had shown that device-based normothermia could reduce fever in patients with severe cerebrovascular disease, without establishing a functional benefit. The PAIS and PAIS 2 trials had not demonstrated improved outcomes with prophylactic paracetamol. QASC supported a broader package addressing fever, glucose and swallowing in stroke units, but could not isolate the contribution of temperature management. These studies left uncertainty about preventive device treatment in critically ill patients.^{2,3,4,5}

INTREPID tested automated surface cooling after ischaemic stroke, intracerebral haemorrhage or subarachnoid haemorrhage, before sustained or recurrent fever had developed. Its clinical ambition extended beyond showing that a machine could control temperature: investigators also sought evidence of improved functional recovery. The trial therefore offered an opportunity to examine both the physiological premise and the patient benefit required to justify routinely adding a substantial intervention to neurocritical care.¹

Synopsis

INTREPID was an open-label, randomised trial with blinded outcome assessment, conducted in 43 intensive care units in seven countries. Recruitment ran from March 2017 to April 2021, with final follow-up in May 2022. Of 686 consented patients, nine were not randomised. The randomised population comprised 677 participants: 339 assigned to fever prevention and 338 to standard care. There were 254 ischaemic strokes, 223 intracerebral haemorrhages and 200 subarachnoid haemorrhages. Median age was 62 years and 51% were women.¹

Patients were aged 18–85 years and expected to require at least 72 hours of intensive care. Baseline modified Rankin scale scores were zero to two, with a score of zero required at ages 81–85. Exclusions included temperature of at least 38°C for more than one hour or on more than one occasion before enrolment, catastrophic injury leading to brain death or comfort-only care, and conditions making six-month survival unlikely. The population was therefore selected for substantial acute neurological illness with a prospect of meaningful recovery.^{1,6}

Disease-specific criteria further defined eligibility. Ischaemic stroke patients entered 3–24 hours after onset and had an NIH Stroke Scale score of at least six. Intracerebral haemorrhage patients entered 6–36 hours after last known well, had haemorrhage volumes of 1–60 mL, an NIH Stroke Scale score of at least six and Glasgow Coma Scale score of at least five. Subarachnoid haemorrhage patients entered within 72 hours of symptoms, with additional requirements concerning imaging and stabilisation after procedures. The different entry windows reflected the distinct clinical pathways.⁶

Central web-based randomisation used minimisation for stroke type, centre, age and disease-specific severity. Fever prevention used the Arctic Sun automated surface device, with adhesive gel pads and core-temperature feedback, targeting 37.0°C for up to 14 days or ICU discharge. Temperature and shivering were assessed hourly. Counterwarming and escalating medication were used to control shivering while attempting to limit sedation. Treating teams knew allocation; trained functional-outcome assessors were masked.^{1,6}

Standard care was an active treatment strategy triggered by temperature of at least 38°C. Escalation began with antipyretic medication, followed by physical measures such as ice packs or fans, then cooling blankets and, if necessary, advanced temperature-management devices. Treatment details retained some clinical discretion. This comparator allowed rescue device treatment, rather than prohibiting effective control of established fever.^{1,6}

The primary endpoint was daily mean fever burden: the area under the temperature curve above 37.9°C, divided by observed acute-phase hours and multiplied by 24. Its unit was °C-hour per day; it was not a difference in average body temperature. The principal secondary endpoint was functional recovery at 90 days, assessed by an ordinal modified Rankin analysis that combined scores five and six into one category. A further analysis classified scores zero to three as favourable and four to six as unfavourable. Other outcomes included later function, cognition, mortality and resource use.^{1,6}

The planned sample of 1,176 allowed for 15% loss and provided 88% power to detect an increase in favourable functional outcome from 40% to 50%. Power for fever burden exceeded 99%. The reported testing sequence considered fever burden before the functional endpoint. A planned interim assessment examined conditional power for the binary clinical outcome; a value below 5% supported stopping. The observed conditional power was 1.9%, and enrolment stopped for clinical futility. Analyses of additional secondary outcomes were not adjusted for multiple comparisons.^{1,6}

The device was started in 330 fever-prevention participants. Median recorded use was 93.2 hours, and 164 of 339 participants had a temporary suspension. Thirty-four controls received advanced device treatment after the earlier fever-management tiers. Approximately one-quarter of controls with temperature data never developed fever. Thus, the assigned contrast was appreciable but neither continuous device exposure in every intervention patient nor universal fever in controls.^{1,6}

Primary temperature data were available for 320 intervention patients and 329 controls. Mean daily fever burden was 0.37 versus 0.73°C-hour, with an adjusted difference of -0.35°C-hour (95% confidence interval -0.51 to -0.20; $P < 0.001$). Fever occurred in 175 of 320 (54.7%) versus 248 of 329 (75.4%), risk ratio 0.73 (0.65–0.82). Total fever duration averaged 9.0 versus 21.6 hours. The reduction in mean fever burden was smaller in ischaemic stroke than in either haemorrhagic subtype, although this alone does not establish a treatment interaction.^{1,6}

Ninety-day Rankin outcomes were available for 278 intervention patients and 306 controls. The ordinal comparison did not demonstrate a difference: reported odds ratio 1.09 (95% confidence interval 0.81–1.46; $P = 0.54$). Scores zero to three occurred in 109 of 278 (39.2%) versus 131 of 306 (42.8%), risk ratio 0.92 (0.76–1.12). Six- and twelve-month functional comparisons likewise did not establish benefit. Multiple-imputation analyses supported the neutral functional result, while an extreme worst-case assignment of missing outcomes favoured control.^{1,6}

Safety analyses used treatment received and therefore different denominators. Adjudicated major adverse events through three months occurred in 157 of 331 device-treated patients (47.4%) and 151 of 336 controls (44.9%). Any acute-phase adverse event occurred in 82.2% versus 75.9%, and serious adverse events in 45.3% versus 39.6%. Shivering was recorded in 283 of 331 (85.5%) versus 82 of 337 (24.3%); 26 fever-prevention patients stopped treatment because of shivering. By twelve months, deaths in the safety population numbered 73 versus 71. The trial did not demonstrate improved survival.^{1,6}

Critique

INTREPID provides persuasive evidence that preventive automated cooling can reduce fever burden in this population. Allocation was randomised centrally, baseline disease severity was incorporated into assignment, temperature was repeatedly measured and the direction of the primary result persisted in sensitivity analyses. The trial also included several vascular diagnoses and many centres, making the physiological finding more informative than a small, single-centre demonstration of device performance. Blinded assessment of later disability strengthened the clinical comparison.^{1,6}

The primary success should nevertheless be kept separate from the therapeutic question. A reduction of 0.35°C-hour in daily fever burden describes less exposure above a threshold; it does not quantify preserved brain function. Both the binary and ordinal clinical analyses were neutral. This does not prove that fever is harmless or entirely a marker of injury, because temperature reduction might be insufficient, mistimed or accompanied by countervailing effects. It does mean that the tested strategy did not establish the patient benefit needed to justify routine preventive use.¹

The principal clinical endpoint deserves more weight than an incidental secondary measurement. The investigators planned recruitment around functional recovery, included it in the hierarchical sequence and used clinical futility to stop enrolment. However, the assumed ten-percentage-point absolute benefit was substantial. Stopping with 677 randomised participants reduced precision relative to the planned programme. The confidence interval remains compatible with smaller effects in either direction; the trial is neither an equivalence study nor evidence that every possible fever-prevention strategy is ineffective.^{1,6}

The 1.9% conditional-power estimate is also easy to overinterpret. It represented the projected chance of reaching the specified statistical objective if the observed interim pattern continued under the calculation used. It was not a 1.9% probability that fever prevention has any clinical benefit. The stopping decision was reasonable within the

reported design, but its meaning remains tied to the targeted effect, interim data and endpoint. It does not settle whether a different population or a more selective intervention could benefit.⁶

Missing functional outcomes are a material limitation. At 90 days, 61 intervention participants and 32 controls lacked Rankin outcomes, leaving unequal ascertainment of 82.0% and 90.5%. Multiple imputation did not reveal a benefit, which is reassuring within its assumptions. Conversely, the appendix's deliberately extreme assignment of poor outcomes to missing intervention patients and good outcomes to missing controls produced a significant control advantage. That scenario is not a credible estimate of harm by itself, but it shows why missingness cannot be dismissed simply because the main imputation analysis was neutral.^{1,6}

Outcome construction also affects clinical interpretation. Combining Rankin scores five and six treats the most severe dependency and death as a single category in the ordinal analysis. The binary definition of favourable outcome included score three, which permits some need for assistance and should not be described as functional independence. Separate survival data and the full distribution therefore remain important. The absence of an established benefit across these measures is more informative than the primary fever endpoint alone.^{1,6}

Exposure was less uniform than the maximum fourteen-day treatment window suggests. Median device use was under four days, almost half the intervention group had a temporary suspension, and some controls received rescue devices. These features could reduce separation, but they also describe how the strategy operated in clinical practice. An analysis restricted to fully adherent patients would sacrifice the protection of randomisation. Moreover, later temperature curves represent changing groups as patients die, leave intensive care or cease observation; visual convergence cannot by itself explain the functional result.^{1,6}

Daily fever burden normalised exposure to the period observed. This made patients with different ICU durations comparable on a rate-like measure, but did not establish that every hour of fever carried the same neurological consequence. The disease-specific entry windows likewise meant that the intervention did not test a single interval from injury across all three diagnoses. A smaller mean temperature effect in ischaemic stroke may reflect different fever patterns or other factors; it is insufficient evidence for a disease-specific treatment recommendation.^{1,6}

Shivering was a substantial treatment burden, despite a normothermic target. It required surveillance, counterwarming and sometimes medication, and caused treatment

cessation in a minority. The high incidence makes it inappropriate to describe prophylactic cooling as a trivial addition to care. It is plausible that associated metabolic stress or treatment could influence recovery, but the trial cannot establish that these effects cancelled a neurological benefit. Safety comparisons were additionally based on treatment received, so they lack the same causal protection as the randomised efficacy analysis.^{1,6}

The appendix reported a small nominal increase in mean hospital stay, while many other resource and cognitive outcomes showed no clear difference. These numerous secondary comparisons were not multiplicity-adjusted, and cognitive assessments were available in a much smaller subset than the randomised population. Neither a definitive resource penalty nor a cognitive benefit should be inferred. Likewise, “any”, “serious” and adjudicated “major” adverse events are different categories; separating them is necessary to avoid overstating or understating the safety findings.⁶

Industry funding warrants transparent consideration. Becton Dickinson supplied support and devices and participated in design, conduct, data management, analysis and interpretation, while the report describes investigator data access and no sponsor role in manuscript preparation or the publication decision. These arrangements do not invalidate the findings, but make the reported masking, independent assessments and complete presentation of the neutral clinical result important. Overall, INTREPID supports temperature efficacy while leaving routine clinical utility unproven.¹

Body of Evidence

The earlier bicentre trial by Broessner and colleagues randomised 102 patients with severe cerebrovascular disease to endovascular normothermia at 36.5°C or stepwise conventional treatment. Median total fever burden was zero versus 4.3°C-hours. Mortality and neurological outcomes did not differ significantly. Its temperature endpoint was cumulative, unlike INTREPID's daily normalised measure, and it used an intravascular intervention. The two studies therefore cannot be compared by subtracting their fever-burden values. They do, however, support the same general observation: successful temperature control does not automatically establish improved neurological recovery.²

PAIS approached the problem pharmacologically. It randomised 1,400 patients with acute ischaemic stroke or intracerebral haemorrhage to paracetamol 6 g daily or placebo within twelve hours. The primary sliding-dichotomy functional outcome did not establish benefit: adjusted odds ratio 1.20 (95% confidence interval 0.96–1.50). A post hoc signal among patients with baseline temperature of 37–39°C suggested a possible role for

temperature-based selection, but could not justify routine treatment. Its broader stroke population and drug-based intervention differed from INTREPID's ICU device strategy.³

PAIS 2 attempted further evaluation but stopped after enrolling 256 of a planned 1,500 patients because recruitment was slow and funding ended. The adjusted common odds ratio for functional outcome was 1.15 (95% confidence interval 0.74–1.79). This imprecise result neither confirmed benefit nor tightly excluded it. Read alongside INTREPID, these trials weaken the assumption that prophylaxis directed at temperature will necessarily improve disability, while leaving uncertainty about which patients, timing and treatment intensity might matter. They do not support transferring the effect estimate for one intervention directly to another.^{1,4}

QASC addressed a different level of care. Nineteen acute stroke units were cluster-randomised to implementation of a package covering fever, hyperglycaemia and swallowing dysfunction, or to guideline dissemination. The study included 687 patients before implementation and 1,009 afterwards. In the post-intervention comparison, death or dependency defined as Rankin score at least two occurred in 42% versus 58%, with an adjusted absolute difference of 15.7 percentage points. This finding supported organised delivery of the combined care package, but did not isolate fever treatment from glucose management, swallowing care, staff education or teamwork.⁵

There is consequently no necessary conflict between QASC and INTREPID. A beneficial multidisciplinary package does not prove that adding preventive cooling to active fever treatment will improve outcomes. Their settings, patient severity, interventions and Rankin definitions were different. Conversely, INTREPID's neutral functional comparison does not justify withdrawing the other elements of organised stroke care or withholding treatment once fever develops. The relevant decision is narrower: whether to begin a demanding device strategy before established fever in patients resembling those enrolled.^{1,5}

The accompanying editorial welcomed the feasibility demonstrated by INTREPID while arguing for more selective future investigation. That is a reasonable research direction, particularly because ischaemic stroke and the two haemorrhagic conditions have different mechanisms and trajectories. Nevertheless, selecting a promising subgroup after observing a neutral trial cannot establish efficacy within it. A future enriched trial would need to define that population prospectively and retain a patient-centred endpoint, rather than treating improved temperature curves as sufficient evidence.⁷

By the end of 2024, INTREPID had strengthened the evidence that automated normothermia prevents fever without demonstrating better functional recovery. Its

findings favour restraint in adopting universal prophylactic device cooling for neuroprotection. They leave a separate role for studying targeted use and for treating refractory fever when temperature control itself is the immediate objective. Such use should not be represented as a proven means of improving long-term neurological outcome on the basis of this trial.¹

Should automated surface cooling be used routinely to prevent fever after severe vascular brain injury?

INTREPID reduced fever burden but did not demonstrate better functional recovery. The findings do not support routine preventive device cooling for neuroprotection in this population.

References

1. Greer DM, Helbok R, Badjatia N, Ko SB, Guanci MM, Sheth KN; INTREPID Study Group. Fever Prevention in Patients With Acute Vascular Brain Injury: The INTREPID Randomized Clinical Trial. *JAMA* 2024;332(18):1525–1534.
2. Broessner G, Beer R, Lackner P, Helbok R, Fischer M, Pfausler B, et al. Prophylactic, endovascularly based, long-term normothermia in ICU patients with severe cerebrovascular disease: bicenter prospective, randomized trial. *Stroke* 2009;40(12):e657–e665.
3. den Hertog HM, van der Worp HB, van Gemert HMA, 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.
4. de Ridder IR, den Hertog HM, van Gemert HMA, Schreuder AHCLT, Ruitenberg A, Maasland EL, et al. PAIS 2 (Paracetamol [Acetaminophen] in Stroke 2): Results of a Randomized, Double-Blind Placebo-Controlled Clinical Trial. *Stroke* 2017;48(4):977–982.

- 
5. Middleton S, McElduff P, Ward J, Grimshaw JM, Dale S, D'Este C, et al. Implementation of evidence-based treatment protocols to manage fever, hyperglycaemia, and swallowing dysfunction in acute stroke (QASC): a cluster randomised controlled trial. *Lancet* 2011;378(9804):1699–1706.
 6. INTREPID Study Group. Supplementary online content to: Fever Prevention in Patients With Acute Vascular Brain Injury: The INTREPID Randomized Clinical Trial. *JAMA*. doi:10.1001/jama.2024.14745.
 7. May T, Shutter L. Feasibility of Fever Prevention in Vascular Brain Injury. *JAMA* 2024;332(18):1521–1522.

Possible Reporting Inconsistencies. The abstract labels 82.2% versus 75.9% as 'major adverse events'. Supplementary eTables 23 and 28 identify these as any acute-phase adverse event; Table 3 reports adjudicated major events separately (47.4% versus 44.9% through three months). These categories should not be conflated.^{1,6}

PROPHY-VAP

Dahyot-Fizelier C, Lasocki S, Kerforne T, Perrigault PF, Geeraerts T, Asehnoune K, et al.; PROPHY-VAP Study Group and ATLANREA Study Group. Ceftriaxone to prevent early ventilator-associated pneumonia in patients with acute brain injury: a multicentre, randomised, double-blind, placebo-controlled, assessor-masked superiority trial. *Lancet Respir Med* 2024;12(5):375–385.

Introduction

Patients intubated after severe brain injury are particularly vulnerable to pneumonia. Loss of airway protection, aspiration around the time of injury and subsequent mechanical ventilation create opportunities for infection. Brain injury may also impair immune responses. Pneumonia can then add respiratory failure, fever and prolonged intensive care to an already serious neurological illness. The clinical attraction of prevention is therefore substantial, but so is the need to distinguish preventable infection from the pulmonary consequences of the initial injury.¹

Giving an antibiotic shortly after intubation could interrupt progression from airway contamination to established pneumonia. Earlier randomised studies of cefuroxime and ampicillin–sulbactam had reported fewer early infections in comatose patients, without establishing broader patient benefit. Their limited size and open treatment designs left uncertainty. ANTHARTIC subsequently demonstrated less early pneumonia after cardiac arrest with a two-day antibiotic course, but did not demonstrate improvements in ventilation duration or survival.^{2,3,4}

Antibiotic prophylaxis has potential benefits and harms. Preventive antibiotics expose patients who might never become infected to treatment and may alter individual or unit-wide bacterial ecology. Conversely, preventing pneumonia could avert longer subsequent antibiotic courses. A short intervention might therefore increase antibiotic exposure immediately while reducing it over the admission. Neither its simplicity nor its brevity can establish the balance between these effects.^{1,5}

PROPHY-VAP tested a single dose of ceftriaxone within twelve hours of intubation in adults with severe acute brain injury. It deliberately evaluated systemic prophylaxis without selective digestive or oropharyngeal decontamination. The trial asked whether a minimal antibiotic exposure could prevent early ventilator-associated pneumonia, while assessing whether any reduction was accompanied by improved patient outcomes and acceptable short-term safety.¹

Synopsis

PROPHY-VAP was a multicentre, randomised, double-blind, placebo-controlled superiority trial in nine intensive care units at eight French university hospitals. Between October 2015 and May 2020, 2,230 patients were screened and 345 were randomised: 171 to ceftriaxone and 174 to placebo. Of these, 330 received the assigned intervention. After post-randomisation exclusions and consent withdrawals, the modified intention-to-treat analysis contained 319 patients, with 162 in the ceftriaxone group and 157 in the placebo group.¹

Adults were eligible after traumatic brain injury, stroke or subarachnoid haemorrhage if their Glasgow Coma Scale score was twelve or lower and mechanical ventilation was expected to continue for more than 48 hours. Randomisation had to occur within twelve hours of tracheal intubation and within 48 hours of hospital admission. Cardiac arrest, tumour-related or infectious coma, imminent death, ongoing antibiotics, recent hospitalisation and anticipated antibiotic prophylaxis during the next 24 hours were among the exclusions. Nasal intubation, tracheostomy, subglottic secretion drainage and contraindications to beta-lactams also excluded participation.¹

Central web-based allocation used variable blocks and stratification by centre and Glasgow Coma Scale below eight versus eight or higher. Participants, clinical teams, outcome assessors and the statistician were masked. An uninvolved nurse from a neighbouring unit prepared the infusion, using opaque syringes and tubing to conceal ceftriaxone's colour. The intervention was ceftriaxone 2 g intravenously over thirty minutes, once; the comparator was saline. Mean administration time was seven hours after intubation in both groups.¹

Both groups received the units' usual pneumonia-prevention measures, including head elevation, cuff-pressure monitoring, oral care and efforts to minimise unnecessary ventilation. Three centres used chlorhexidine mouth care and six used sterile water. Adherence to the overall prevention bundle was not formally monitored. No centre used selective digestive or oropharyngeal decontamination. Subsequent suspected infections were investigated and treated according to local practice based on national guidance.¹

The primary outcome was adjudicated early ventilator-associated pneumonia, defined in the Methods as occurring from the second through seventh day of mechanical ventilation, with diagnostic criteria requiring at least 48 hours of ventilation. Diagnosis required clinical, radiological and microbiological evidence. At least two clinical features were needed: abnormal temperature, abnormal white-cell count or purulent secretions. Quantitative culture thresholds depended on sampling method: 10^3 colony-forming units/mL for protected distal sampling, 10^4 for bronchoalveolar lavage and 10^6 for

endotracheal aspirate. Two masked senior intensivists reviewed suspected cases, with a third resolving disagreement.¹

Secondary outcomes included later pneumonia, all pneumonia through day 28, ventilator-associated events, survival, neurological function, adverse events and days alive without ventilation or antibiotics. ICU-free and hospital-free days were measured through day 60. Death before the relevant horizon assigned zero event-free days. Four units routinely collected admission and discharge rectal swabs for extended-spectrum beta-lactamase-producing organisms; this was a limited surveillance component rather than a trial-wide microbiome assessment.^{1,6}

The sample calculation required 354 evaluable patients to detect a reduction in early pneumonia from 30% to 15%, with 90% power and a two-sided 5% significance level. The published analysis used an adjusted Fine–Gray competing-risk model, treating death before pneumonia as a competing event and ventilator weaning as censoring. The report describes changing from an initially planned Cox model to this approach. Other continuous outcomes were compared using parametric or rank-based tests as appropriate. Missing outcomes were not imputed.¹

The analysed population had a mean age of 57 years and was 48% female. Subarachnoid haemorrhage was the largest diagnostic category, followed by traumatic brain injury and haemorrhagic stroke; ischaemic stroke accounted for fewer than one in ten participants. Approximately three-quarters had Glasgow Coma Scale scores of eight or lower. Subsequent surgical or procedural interventions were common. Despite excluding anticipated early prophylaxis, 24 patients later received surgical antibiotic prophylaxis, with similar numbers in the two groups.^{1,6}

Investigators reported 160 suspected pneumonia episodes in 139 patients. Adjudication confirmed 93 episodes in 90 patients, including 74 early infections. Early pneumonia occurred in 23 of 162 ceftriaxone patients (14.2%) and 51 of 157 controls (32.5%): reported adjusted hazard ratio 0.60 (95% confidence interval 0.38–0.95; $P=0.030$). The crude absolute difference, calculated from these proportions, was 18.3 percentage points. This risk difference is a separate summary from the model-based hazard ratio.¹

By day 28, at least one adjudicated pneumonia had occurred in 33 ceftriaxone patients and 57 controls, with 35 versus 58 episodes: reported hazard ratio 0.62 (0.42–0.98). Late pneumonia affected eleven versus seven patients. Median ventilator-free days were nine versus five, and antibiotic-free days 21 versus fifteen. Twenty-eight-day mortality was 25 of 162 (15%) versus 39 of 157 (25%), reported hazard ratio 0.62 (0.39–0.97; $P=0.036$).

Investigators attributed 59 of these 64 deaths to consequences of the initial brain injury.¹

At day 60, median ICU-free days were 34 versus 26 and hospital-free days 23 versus eight. Mortality was 32 of 161 versus 46 of 157, hazard ratio 0.66 (0.42–1.04; $P=0.074$). The Rankin distribution favoured ceftriaxone at day 28 in a secondary comparison, but did not differ significantly at day 60. The report recorded 194 adverse events in 152 patients. Resistant respiratory isolates were uncommon. Among 115 patients reported to have rectal-swab surveillance, two acquired extended-spectrum beta-lactamase-producing organisms, both in the ceftriaxone group.^{1,6}

Critique

The trial substantially improves the evidence for a specific, practicable intervention. Central randomisation, a saline comparator, concealed infusion appearance and masked clinical assessment address several weaknesses of earlier studies. The dose and timing are straightforward, and the participating units treated a range of structural brain injuries. Public funding, with no reported funder role in design, analysis or writing, further supports the trial's independence from commercial interests. The question concerned adding one dose to an existing prevention bundle, not replacing basic infection-prevention care.¹

The primary result is clinically sizeable, but its endpoint requires care. Antibiotics can lower bacterial counts in respiratory samples as well as prevent invasive infection. Because culture positivity above a threshold was required, ceftriaxone could affect whether a suspected episode met the definition even if some pulmonary illness remained. Masked adjudication reduces subjective differential classification but cannot remove this biological effect on the diagnostic test. The accompanying editorial identifies this distinction as a central issue in interpreting antibiotic-prophylaxis trials.^{1,5}

Adjudication nevertheless added substantial value. Forty-two per cent of initially suspected episodes were not confirmed, commonly because the microbiological threshold or a new radiographic infiltrate was absent. This demonstrates how different a clinical suspicion of pneumonia can be from a standardised trial endpoint. Sampling approaches were similarly distributed between groups, making a simple imbalance in proximal versus distal sampling an unlikely explanation. The adjudicators were members of the study team rather than an external committee, but were masked to assignment.¹

The wider pattern of results is encouraging because it extends beyond cultures. Fewer deaths at day 28 and more days alive without ventilation, antibiotics or hospitalisation

are compatible with genuine patient benefit. However, these are secondary outcomes, the sample size was calculated for pneumonia and many comparisons were made. Their coherence adds plausibility without making each result independently definitive. The mortality confidence interval at day 60 crossed the null, and no sustained functional advantage was established. A survival benefit should therefore remain a promising finding requiring confirmation.¹

Event-free days are especially important to interpret correctly. A patient dying before the horizon received zero days, so the observed differences combine survival with freedom from treatment or hospital care. More antibiotic-free days cannot be translated directly into six fewer treatment days per patient, nor do more hospital-free days necessarily mean that surviving patients left seventeen days sooner. These measures appropriately give poor outcomes to deaths, but they do not isolate the mechanism of improvement. The appendix's raw lengths of stay were much closer than the corresponding free-day medians.^{1,6}

The primary hazard ratio is likewise not the crude relative risk implied by 14% versus 32%. It comes from a model adjusted for stratification variables and accounting for death as a competing event. Weaning was censored, although stopping ventilation removes or changes the opportunity to develop ventilator-associated pneumonia. Interpretation therefore depends partly on how treatment affects both infection and remaining at risk. The reported change from Cox to competing-risk analysis is relevant context, but the main publication alone does not establish its timing relative to access to outcomes.¹

Post-randomisation exclusions deserve attention. Fifteen participants did not receive treatment and eleven withdrew consent, leaving nine fewer intervention patients and seventeen fewer controls than originally allocated. Some exclusions reflected legal or consent requirements and others eligibility discovered after allocation. The resulting analysis answers the question in the retained population and does not preserve every randomised participant. It also contained fewer evaluable patients than the planned 354. These limitations temper precision and protection against selection, although the primary effect remained substantial.¹

Applicability is narrower than the abstract's recommendation for all ventilated patients with brain injury. Only 345 of 2,230 screened patients were randomised. The largest screening exclusion was existing or anticipated antibiotic exposure, while delayed presentation beyond twelve hours of ventilation excluded another large group. Units treating many patients already receiving antibiotics, patients with recent hospital contact or substantial resistant colonisation may obtain a different balance of benefit

and harm. The trial also provides no direct comparison in units using subglottic drainage or selective digestive decontamination.¹

Resistance surveillance offers limited reassurance. Only a subset underwent rectal sampling, respiratory resistant isolates were rare and follow-up was short. Two acquisitions in the ceftriaxone group versus none in controls are too few to establish either an excess risk or ecological safety. An individually randomised study in mixed-treatment units cannot determine what happens to unit-wide resistance after repeated, sustained adoption. The absence of an obvious short-term signal should therefore not be described as proof of no microbiological impact.^{1,5}

The data also cannot firmly establish how ceftriaxone affected every subsequent antibiotic decision. The appendix lists prescribed agents, but does not provide a comprehensive longitudinal account of spectrum, indication and treatment duration for all later infections. Non-ventilated pneumonia and other hospital-acquired infections were not systematically evaluated as efficacy outcomes. These omissions matter when proposing an overall antibiotic-saving policy. Similarly, a baseline chest radiograph was not mandatory, so some recipients may already have had incipient infection rather than being entirely uninfected at the time of prophylaxis.^{1,6}

Finally, the neurological outcomes were measured relatively early and had missing assessments, particularly at day 28. Most deaths were attributed to the original brain injury, and the trial did not measure the proposed intermediate mechanism of fewer secondary cerebral insults. Preventing pneumonia could plausibly help recovery, but this causal chain was not demonstrated. PROPHY-VAP is strong evidence for less adjudicated early pneumonia in a selected population, with useful supportive clinical signals; it is less conclusive about long-term recovery, survival and the consequences of routine widespread antibiotic use.¹

Body of Evidence

Sirvent and colleagues provided an early randomised comparison in 100 ventilated patients with head injury or stroke and Glasgow Coma Scale scores of twelve or lower. Two 1.5-g doses of cefuroxime, twelve hours apart after intubation, were associated with microbiologically confirmed pneumonia in 24% versus 50%. Mortality and morbidity did not differ significantly. The trial was open, and seventeen controls received prophylactic antibiotics at clinicians' discretion. It supported biological efficacy but left uncertainty about the size and clinical consequences of the effect.²

Acquarolo and colleagues subsequently compared three days of ampicillin–sulbactam with standard treatment in comatose ventilated patients. This small, single-centre, open trial stopped after an interim assessment showed less early pneumonia: 21.0% versus 57.9%. Other outcomes did not establish a benefit, and the investigators explicitly recognised the limits imposed by sample size. An early positive stopping decision in a small study also makes the magnitude of benefit uncertain. PROPHY-VAP's masked, multicentre design and single-dose regimen therefore represent an important advance over this evidence.^{1,3}

ANTHARTIC tested a two-day course of intravenous amoxicillin–clavulanate after out-of-hospital cardiac arrest with an initially shockable rhythm, in patients receiving temperature management at 32–34°C. Of 198 randomised patients, 194 entered the analysis. Adjudicated early pneumonia occurred in 19% versus 34%, hazard ratio 0.53 (95% confidence interval 0.31–0.92). Ventilator-free days and day-28 mortality did not differ significantly. This supports the possibility of preventing early pneumonia after a neurological insult, while illustrating that an infection endpoint and a survival benefit need not move together.⁴

Cardiac-arrest patients were excluded from PROPHY-VAP, so the two trials address related but different populations. Their neurological prognosis, temperature treatment, antibiotic regimens and timing differed. The broadly consistent pneumonia results are informative, but combining them conceptually cannot establish that one ceftriaxone dose is equivalent to two days of amoxicillin–clavulanate. Neither trial answers every question about prophylaxis in an undifferentiated ICU population.^{1,4}

PASS provides a further caution against extrapolation from ventilated brain-injury patients to all patients with stroke. In this Dutch trial, 2,550 patients were randomised to ceftriaxone 2 g daily for four days or standard stroke-unit care, beginning within 24 hours of symptoms. After twelve immediate consent withdrawals, 2,538 entered analysis. The three-month Rankin distribution was not improved: adjusted common odds ratio 0.95 (95% confidence interval 0.82–1.09; $P=0.46$). The same antibiotic, given longer, did not establish functional benefit in that broader stroke setting.⁷

These comparisons suggest that patient selection and the treatment objective are central. PROPHY-VAP concentrated on patients with impaired consciousness and prolonged invasive ventilation, where early pneumonia risk was high. PASS tested functional recovery across a broader stroke population. Selective digestive decontamination involves another intervention altogether, combining repeated topical treatment with systemic antibiotics; PROPHY-VAP did not compare its single dose against that package. Its result therefore supports neither replacing an established

decontamination programme nor assuming that the strategies have identical ecological consequences.^{1,5,7}

By the end of 2024, the evidence for reducing early pneumonia with short systemic prophylaxis in selected comatose patients was considerably stronger. PROPHY-VAP added promising clinical signals and a notably simple regimen. A reasonable appraisal is to consider this strategy in closely matching patients alongside local microbiology and existing prevention practice, while acknowledging that the authors' call for universal adoption goes beyond what is securely established about mortality and resistance. Larger and longer assessments of those outcomes would materially strengthen the case for routine implementation.^{1,5}

Does one early dose of ceftriaxone prevent pneumonia after severe brain injury requiring ventilation?

A single 2-g dose within twelve hours of intubation reduced adjudicated early pneumonia in selected patients. Survival benefit and the ecological safety of widespread use remain uncertain.

References

1. Dahyot-Fizelier C, Lasocki S, Kerforne T, Perrigault PF, Geeraerts T, Asehnoune K, et al.; PROPHY-VAP Study Group and ATLANREA Study Group. Ceftriaxone to prevent early ventilator-associated pneumonia in patients with acute brain injury: a multicentre, randomised, double-blind, placebo-controlled, assessor-masked superiority trial. *Lancet Respir Med* 2024;12(5):375–385.
2. Sirvent JM, Torres A, El-Ebiary M, Castro P, de Batlle J, Bonet A. Protective effect of intravenously administered cefuroxime against nosocomial pneumonia in patients with structural coma. *Am J Respir Crit Care Med* 1997;155(5):1729–1734.
3. Acquarolo A, Urli T, Perone G, Giannotti C, Candiani A, Latronico N. Antibiotic prophylaxis of early onset pneumonia in critically ill comatose patients. A randomized study. *Intensive Care Med* 2005;31(4):510–516.
4. François B, Cariou A, Clere-Jehl R, Dequin PF, Renon-Carron F, Daix T, et al.; CRICS-TRIGGERSEP Network and ANTHARTIC Study Group. Prevention of Early Ventilator-Associated Pneumonia after Cardiac Arrest. *N Engl J Med* 2019;381(19):1831–1842.

5. Klompas M. Antibiotic prophylaxis for patients with acute brain injury. *Lancet Respir Med* 2024;12(5):340–341.
6. PROPHY-VAP Study Group. Supplementary appendix 2 to: Ceftriaxone to prevent early ventilator-associated pneumonia in patients with acute brain injury. *Lancet Respir Med*. doi:10.1016/S2213-2600(23)00471-X.
7. Westendorp WF, Vermeij JD, Zock E, Hooijenga IJ, Kruijt ND, Bosboom HJLW, et al. The Preventive Antibiotics in Stroke Study (PASS): a pragmatic randomised open-label masked endpoint clinical trial. *Lancet* 2015;385(9977):1519–1526.

Possible Reporting Inconsistencies. The Results text places the suspected treatment-related anaphylactic event in the placebo group, whereas Table 4 lists one anaphylactic shock in the ceftriaxone group and none with placebo. The allocation is therefore unclear, and the statement that no adverse effect was attributable to ceftriaxone should be interpreted cautiously.¹

LASTE

Costalat V, Jovin TG, Albucher JF, Cognard C, Henon H, Nouri N, et al.; LASTE Trial Investigators. Trial of Thrombectomy for Stroke with a Large Infarct of Unrestricted Size. *N Engl J Med* 2024;390(18):1677–1689.

Introduction

A large established infarct was traditionally a reason to hesitate before thrombectomy. Reopening an occluded artery might preserve remaining viable tissue, but could also cause haemorrhage in damaged brain. If little useful tissue remained, successful reperfusion might change neither disability nor survival. Imaging was consequently used to identify patients with a favourable balance between threatened and already injured tissue, leaving those with extensive early changes poorly represented in the original thrombectomy evidence.¹

By the time LASTE reported, several randomised trials had challenged this restrictive approach. RESCUE-Japan LIMIT, SELECT2, ANGEL-ASPECT and TENSION demonstrated functional benefit in selected patients with large infarcts. Their eligibility criteria, imaging methods and treatment windows differed, however, and considerable uncertainty remained at the most severe end of the imaging spectrum. A positive result in a broadly defined large-core population could not establish that every patient with extensive injury would benefit.^{2,3,4,5}

LASTE was designed to examine this boundary. It allowed an Alberta Stroke Program Early Computed Tomography Score (ASPECTS) as low as zero in younger participants and imposed no upper infarct-volume threshold. The central question was whether early thrombectomy could still improve patient-important outcomes when substantial injury was already visible. Its title's reference to unrestricted size needs to be understood within the trial's other selection criteria: these were previously independent patients with proximal anterior-circulation occlusion, and established mass effect with midline shift remained an exclusion.^{1,6}

For intensive care practice, the distinction between less disability, independence and survival is particularly important. A treatment may produce a worthwhile shift in outcome without returning most patients to independent living. LASTE therefore needs to be read both as a test of reperfusion in extensive stroke and as evidence for the conversations that follow about likely survival, dependence and the limits of recovery.¹

Synopsis

LASTE was an investigator-initiated, randomised, open-label trial with masked outcome assessment at thirty hospitals in France and Spain. Adults with acute ischaemic stroke due to proximal anterior-circulation large-vessel occlusion were assigned to thrombectomy plus medical care or medical care alone. The trial randomised 333 patients, 166 and 167 respectively. Nine were subsequently excluded for consent or legal reasons, leaving 324 in the principal analysis: 159 assigned to thrombectomy and 165 to medical care.¹

Eligibility required a National Institutes of Health Stroke Scale score of at least six, a prestroke modified Rankin scale score of zero or one, and ASPECTS zero to five on non-contrast CT or diffusion-weighted MRI. ASPECTS ranges from zero to ten, with lower values indicating more extensive injury. For patients aged eighty or older, the supplementary eligibility criteria restricted ASPECTS to four or five. Important exclusions included intracranial haemorrhage, significant mass effect with midline shift, occlusions involving multiple vascular territories and a life expectancy below six months.^{1,6}

Patients with known onset had to undergo randomisation within 6.5 hours of last known well, with arterial puncture intended within a further thirty minutes. Those with unknown onset could enter within 24 hours of last known well if diffusion-weighted MRI showed an acute lesion without a corresponding positive fluid-attenuated inversion recovery signal. This was a tissue-timing criterion, rather than a requirement for a favourable perfusion mismatch. Advanced perfusion imaging was optional, and no maximum core volume was specified.^{1,6}

Central web-based randomisation used minimisation across important prognostic factors, including age, stroke severity, ASPECTS, occlusion location, treatment timing, thrombolysis and transfer status. Clinical teams could not be masked to an invasive procedure. Certified assessors masked to allocation evaluated functional outcomes at ninety and 180 days, while an independent core laboratory reviewed imaging. Clinical events were independently adjudicated, although procedural complications were necessarily assessed with knowledge of treatment.^{1,6}

Operators could use approved thrombectomy devices and their preferred technique. Both groups received guideline-based medical management, with intravenous thrombolysis at the treating clinician's discretion. Of the 159 patients assigned to intervention, 151 underwent thrombectomy; eight had spontaneous recanalisation before device treatment. One control underwent angiography but did not receive thrombectomy. Combined aspiration and stent-retriever treatment was the most frequent initial approach. Successful reperfusion, defined as modified Thrombolysis in

Cerebral Infarction grade 2b or better, was achieved in 130 of 151 treated patients (86.1%).^{1,6}

The primary outcome was the ninety-day modified Rankin scale distribution, combining scores five and six into one category for analysis. The scale otherwise runs from no symptoms at zero to death at six; five indicates severe dependence requiring constant care. The planned proportional-odds approach was unsuitable because its assumption was not met. The reported primary comparison therefore used a Wilcoxon–Mann–Whitney analysis, expressed as a generalised odds ratio, with higher values favouring thrombectomy. Missing ninety-day functional outcomes in one patient per group were singly imputed.^{1,6}

Secondary outcomes included independence, defined as Rankin zero to two; the ability to walk without another person's assistance, represented by zero to three; utility-weighted Rankin scores; quality of life; early neurological improvement; infarct growth; and outcomes at 180 days. Safety assessments included mortality, symptomatic intracranial haemorrhage and procedural complications. The publication explicitly states that confidence intervals for secondary outcomes were not adjusted for multiple comparisons.¹

The intended sample was 450 patients, providing 80% power for the assumed treatment effect. Recruitment stopped early after other large-infarct trials reported benefit. The monitoring board's supplementary correspondence shows that it considered both the external evidence and an unplanned interim examination of LASTE outcomes available in 298 patients. Investigators did not receive those interim results before follow-up was completed. The stopping decision therefore involved emerging internal evidence as well as the changed external evidence base.^{1,6}

Median age was 74 years, median stroke-severity score 21 and median core-laboratory ASPECTS two. MRI was the selection modality in 271 of 324 patients (83.6%). Median infarct volume was 135 mL, with an interquartile range of 106–185 mL. Approximately 56% had ASPECTS two or lower. Intravenous thrombolysis was given to about 35% in each group. Median time from randomisation to puncture was 27 minutes; median last-known-well to puncture time was 305 minutes. Thirty per cent had an unknown onset time, and some entered beyond the usual 6.5-hour window through the MRI criterion.^{1,6}

The primary functional distribution favoured thrombectomy: generalised odds ratio 1.63 (95% confidence interval 1.29–2.06; $P < 0.001$). The uncollapsed median Rankin score was four with thrombectomy and six with medical care. Among patients with available

ninety-day assessments, independence occurred in 21 of 158 (13.3%) versus eight of 164 (4.9%), adjusted relative risk 2.39 (1.18–6.22). Rankin zero to three occurred in 53 of 158 (33.5%) versus twenty of 164 (12.2%), adjusted relative risk 2.62 (1.72–4.36).¹

Ninety-day mortality was 57 of 158 (36.1%) versus 91 of 164 (55.5%), adjusted relative risk 0.65 (0.50–0.84). The crude absolute difference was 19.4 percentage points. At 180 days, the Rankin distribution continued to favour thrombectomy, generalised odds ratio 1.71 (1.35–2.18), with independence in 29 of 157 (18.5%) versus eight of 162 (4.9%). Early neurological improvement and the utility-weighted functional scores also favoured intervention. Mean infarct growth at 24 hours was 51.6 versus 119.5 mL, while decompressive craniectomy occurred in fourteen versus nineteen patients.^{1,6}

Symptomatic intracranial haemorrhage within 24 hours, using the Heidelberg definition, occurred in fifteen of 157 intervention patients (9.6%) and nine of 157 controls (5.7%): adjusted relative risk 1.73 (0.78–4.68). With the narrower SITS-MOST definition, the figures were five versus four. Any intracranial haemorrhage was more frequent after thrombectomy, at 72.0% versus 60.5%, but included radiographic bleeding without clinical deterioration. Eleven procedure-related complications were reported, including vessel perforation, dissection, access-site injury and embolisation into a new territory.^{1,6}

Critique

LASTE provides persuasive evidence that extensive visible infarction does not inevitably eliminate the benefit of thrombectomy. Its randomised comparison, masked functional assessment, central imaging review and almost complete ninety-day outcome ascertainment support internal validity. The primary result was accompanied by improvements in several clinically coherent outcomes and persisted to six months. The combination of less infarct growth, better function and fewer deaths strengthens the interpretation that reperfusion produced a real benefit in this selected population.^{1,6}

The principal estimate must nevertheless be described accurately. The generalised odds ratio is a summary of the relative ordering of outcomes between groups; it is not the adjusted common odds ratio from a proportional-odds regression. It also does not mean that thrombectomy made independence 63% more likely. The reported primary method accommodates a failure of the proportional-odds assumption, while the separately reported binary outcomes show how often particular levels of recovery were reached. Mixing these estimands would make the benefit harder to interpret.^{1,6}

Combining Rankin five and six gives severe dependence and death the same rank in the primary analysis. This avoids claiming that every transition from death to profound

disability is necessarily a patient-valued gain, but also means the primary result alone cannot describe that trade-off. The full outcome distribution and mortality results are therefore essential. Approximately one-third of intervention recipients could walk without another person's assistance at ninety days, but only 13% were independent. Most remained substantially disabled or had died despite successful treatment.¹

The mortality reduction is large and clinically important. It should be reported alongside, rather than substituted for, the prespecified functional question described in the paper. Mortality and the numerous other secondary outcomes did not have multiplicity-adjusted confidence intervals, and the trial was not separately sized to prove every secondary effect. Nevertheless, a substantial survival difference consistent with the functional distribution adds meaningful support. It would be equally misleading to dismiss it solely because it was not the primary efficacy endpoint or to treat each favourable secondary result as an independent definitive finding.¹

Early stopping affects the certainty of the effect's magnitude. Only 333 of the intended 450 participants were randomised. The ethical context had changed after other positive trials, but the monitoring board also reviewed accumulating LASTE data. Stopping after favourable emerging evidence can leave an effect estimate more optimistic than it would have been after full recruitment. The study's strong primary result and consistency with external trials support the direction of benefit, while the incomplete sample reduces precision for its size and for harms.^{1,6}

The post-randomisation exclusions were few but unequal: seven intervention patients and two controls were removed before the principal analysis. Consent and legal constraints explain why retaining everyone may not have been possible, but the resulting population does not contain every randomised patient. In contrast, only two ninety-day functional outcomes were missing in the retained cohort. Available-case and per-protocol analyses were similar to the primary result, and removing individual centres in sensitivity analyses did not reverse the treatment direction. These checks make a simple missing-data or single-centre explanation less plausible.^{1,6}

The safety findings warrant balanced language. The abstract describes more symptomatic haemorrhage with thrombectomy, but the Heidelberg comparison was imprecise and its confidence interval included no increase. The trial therefore observed numerically more symptomatic bleeding without establishing the size, or even the presence, of an excess under that definition. It certainly did not establish absence of harm. Radiographic haemorrhage, symptomatic haemorrhage and specific severe haematoma categories are different outcomes and should not be combined into one unqualified bleeding claim.¹

“Unrestricted size” is a statement about the absence of an upper volume threshold, not an absence of clinical selection. Significant mass effect with midline shift was excluded, baseline independence was required and the oldest patients could not enter with the lowest ASPECTS. The study consequently offers less direct guidance for established malignant swelling, substantial prior disability or very elderly patients with extensive injury. Those boundaries are more useful clinically than treating the title as permission to disregard every imaging feature.^{1,6}

Imaging modality also limits transferability. More than four-fifths were selected with MRI, and ASPECTS on diffusion-weighted imaging is not simply interchangeable with the same numerical score on CT. Local and central ASPECTS assessments differed, illustrating measurement variability in a rapidly evolving lesion. The much smaller CT subgroup had a wide confidence interval; that does not establish that CT-selected patients fail to benefit. It does mean that LASTE alone is less precise about an exclusively CT-based pathway than about its predominantly MRI-selected population.^{1,6}

The subgroup findings are encouraging but do not identify a universal lower boundary for benefit. Patients with very low ASPECTS and those with volumes above 150 mL had estimates favouring intervention. These comparisons involved smaller samples, multiple examinations and, for volume analyses, post hoc assessment. They cannot establish that all patients with a particular large volume benefit or that treatment effects are identical across severity groups. Infarct location, swelling, timing and the patient's prior function remain relevant features of the decision.^{1,6}

Published correspondence emphasised the risk of extrapolating to destruction of most of a vascular territory. The investigators replied that imaging abnormalities do not map perfectly onto irreversible tissue loss, and proposed salvage of remaining tissue and reduced oedema as possible explanations. These are plausible interpretations rather than mechanisms proven by the trial. LASTE's measured reduction in infarct growth supports a biological effect, but does not reveal how much of the functional gain came from each pathway.^{1,7,8}

Finally, rapid access to experienced centres was part of the intervention's delivery. Median puncture followed randomisation by less than half an hour, and reperfusion was usually successful. Results from this setting cannot quantify the benefit after lengthy additional delays or substantially lower procedural success. The sponsor was Montpellier University Hospital, supported by an unrestricted consortium of device manufacturers, with no reported funder role in design or implementation. Commercial support merits disclosure, while the trial's independent adjudication and investigator access to the data are reassuring safeguards.^{1,6}

Body of Evidence

RESCUE-Japan LIMIT was an early randomised challenge to excluding large infarcts. In 203 patients with ASPECTS three to five, thrombectomy increased the proportion reaching Rankin zero to three at ninety days from 12.7% to 31.0%, relative risk 2.43 (95% confidence interval 1.35–4.37). Selection allowed an MRI-based extension beyond six hours. More radiographic haemorrhage occurred with intervention. Its result established that substantial baseline injury and functional benefit could coexist, while leaving lower ASPECTS values and broader imaging pathways uncertain.²

SELECT2 extended the evidence across an international population treated within 24 hours. Its 352 patients qualified through ASPECTS three to five or an imaging core of at least 50 mL. The trial stopped early for efficacy, with a generalised odds ratio of 1.51 (1.20–1.89) for better ninety-day function. Independence occurred in 20% versus 7%. Its alternative volume criterion is important: the trial should not be described as categorically excluding every patient with ASPECTS below three.³

ANGEL-ASPECT enrolled 456 patients in China within 24 hours, using ASPECTS or volume-based eligibility. It also stopped early for efficacy. The ninety-day functional distribution favoured endovascular treatment, generalised odds ratio 1.37 (1.11–1.69), while symptomatic haemorrhage occurred in 6.1% versus 2.7%. These findings supported expansion of treatment in a different health-care setting, while reinforcing that improved function can coexist with greater bleeding exposure.⁴

TENSION supplied particularly relevant evidence for simpler imaging selection. Among 253 patients with ASPECTS three to five treated within twelve hours, predominantly selected by non-contrast CT, thrombectomy improved the ninety-day Rankin distribution: adjusted common odds ratio 2.58 (1.60–4.15). The trial stopped at its first planned interim assessment. Its result supported benefit without requiring routine advanced perfusion selection, although its imaging severity range remained less extensive than LASTE's.⁵

These studies approached a similar clinical problem through different entry criteria. Their consistent direction supports a real benefit of thrombectomy in appropriately selected large-infarct stroke. It does not make their effect estimates interchangeable. Different imaging modalities, severity distributions, treatment windows, endpoint definitions and stopping decisions can all influence apparent effect size. In particular, a common odds ratio, a generalised odds ratio and a relative risk for a dichotomised Rankin outcome answer related but different statistical questions.^{1,2,3,4,5}

LASTE's distinctive contribution was to include many patients with very low ASPECTS and large measured volumes early after presentation. It therefore reduced an important area of uncertainty left by the earlier trials. Its findings make it harder to justify refusing thrombectomy solely because the initial scan shows extensive injury. At the same time, its age restriction at the lowest scores, exclusion of established mass effect and predominance of MRI prevent a claim that imaging severity and patient selection have become irrelevant.^{1,6}

For clinicians discussing treatment with families, the absolute outcomes remain more informative than the headline relative estimate alone. The chance of independent living was improved but remained low, while a larger group achieved the ability to walk without assistance and mortality was substantially lower. Whether those possibilities align with an individual patient's values cannot be answered by an ordinal score. The trial provides probabilities for that discussion, with considerable uncertainty around any particular patient's trajectory.¹

By the end of 2024, LASTE had strengthened the case for offering prompt thrombectomy to patients matching its criteria despite extensive early infarction. The evidence supports active assessment for treatment rather than assuming futility from a low ASPECTS alone. It also supports honest prognostic counselling: successful reperfusion often reduces disability without restoring independence, and the most extreme presentations outside the trial's eligibility remain an area of uncertainty.^{1,6}

Does thrombectomy help selected patients with a very large acute anterior-circulation infarct?

Early thrombectomy improved functional outcomes and was associated with lower mortality, although most patients did not regain independence. The result does not establish benefit outside the trial's clinical and imaging selection criteria.

References

1. Costalat V, Jovin TG, Albucher JF, Cognard C, Henon H, Nouri N, et al.; LASTE Trial Investigators. Trial of Thrombectomy for Stroke with a Large Infarct of Unrestricted Size. *N Engl J Med* 2024;390(18):1677–1689.
2. Yoshimura S, Sakai N, Yamagami H, Uchida K, Beppu M, Toyoda K, et al.; RESCUE-Japan LIMIT Investigators. Endovascular Therapy for Acute Stroke with a Large Ischemic Region. *N Engl J Med* 2022;386(14):1303–1313.

3. Sarraj A, Hassan AE, Abraham MG, Ortega-Gutierrez S, Kasner SE, Hussain MS, et al.; SELECT2 Investigators. Trial of Endovascular Thrombectomy for Large Ischemic Strokes. *N Engl J Med* 2023;388(14):1259–1271.
4. Huo X, Ma G, Tong X, Zhang X, Pan Y, Nguyen TN, et al.; ANGEL-ASPECT Investigators. Trial of Endovascular Therapy for Acute Ischemic Stroke with Large Infarct. *N Engl J Med* 2023;388(14):1272–1283.
5. Bendszus M, Fiehler J, Subtil F, Bonekamp S, Aamodt AH, Fuentes B, et al.; TENSION Investigators. Endovascular thrombectomy for acute ischaemic stroke with established large infarct: multicentre, open-label, randomised trial. *Lancet* 2023;402:1753–1763.
6. LASTE Trial Investigators. Supplementary appendix to: Trial of Thrombectomy for Stroke with a Large Infarct of Unrestricted Size. *N Engl J Med* 2024. doi:10.1056/NEJMoa2314063.
7. Chen H, Malhotra A, Gandhi D. Thrombectomy for Large Strokes. *N Engl J Med* 2024;391(4):378–379.
8. Arquizan C, Jovin T, Costalat V. Thrombectomy for Large Strokes. Reply. *N Engl J Med* 2024;391(4):379.

TESLA

Yoo AJ, Zaidat OO, Sheth SA, Rai AT, Ortega-Gutierrez S, Given CA II, et al.; Writing Committee for the TESLA Investigators. Thrombectomy for Stroke With Large Infarct on Noncontrast CT: The TESLA Randomized Clinical Trial. JAMA 2024;332(16):1355–1366.

Introduction

The expansion of thrombectomy into large-infarct stroke raised a practical question: how much imaging is necessary before treatment can be offered? MRI and CT perfusion can identify extensive tissue injury while also characterising the surrounding brain. They require equipment, expertise and time that are not equally available across stroke services. Non-contrast CT is more widely accessible, but a large infarct visible on that scan may represent a different stage of injury from a similarly labelled lesion on more sensitive imaging.¹

The distinction matters because the earlier positive large-infarct trials did not all enrol interchangeable populations. Some used predominantly MRI, others combined ASPECTS and perfusion-volume criteria, and treatment windows ranged from early presentation to 24 hours. TENSION supported thrombectomy using mainly non-contrast CT, but within a shorter window. A general recommendation based on these studies still left uncertainty about selecting patients on non-contrast CT alone throughout the full first day.^{1,2}

TESLA addressed that uncertainty in US academic and community hospitals. It accepted patients with ASPECTS two to five, assessed locally on non-contrast CT, and allowed randomisation up to 24 hours after onset or last known well. CT angiography was still needed to identify the occluded artery. “CT alone” therefore describes infarct assessment without mandatory perfusion or MRI selection, rather than a pathway dispensing with vascular imaging.¹

The trial used a Bayesian adaptive design and a utility-weighted functional outcome. This combination makes its interpretation particularly instructive. An estimate can favour treatment and remain compatible with an important benefit while failing the trial's predefined success criterion. TESLA should therefore be examined for what it established, what it left uncertain and how closely its patients resemble those in the other large-infarct trials.^{1,3}

Synopsis

TESLA was a phase 3, randomised, open-label trial with masked endpoint assessment. Forty-seven US centres participated and 44 recruited patients between July 2019 and October 2022. Of 302 patients randomised, two were immediately withdrawn, one following notification of control allocation and one because consent had been obtained from the wrong patient. The retained cohort comprised 300 patients: 152 assigned to thrombectomy plus usual medical care and 148 to medical care alone.¹

Patients were eligible at ages eighteen to 85 years with a National Institutes of Health Stroke Scale score of at least six and prestroke modified Rankin scale zero or one. CT angiography had to show intracranial internal carotid or M1 middle cerebral artery occlusion. A large infarct was defined by non-contrast CT ASPECTS two to five, with lower scores indicating more extensive changes. Randomisation had to occur within 24 hours of symptom onset or last known well. Intracranial haemorrhage, mass effect with ventricular effacement or midline shift, herniation and tandem cervical carotid occlusion were among the exclusions.^{1,4}

Local investigators underwent ASPECTS training and certification. Automated scoring software was offered to sites but could not be used to determine eligibility. Central internet allocation used variable permuted blocks, stratified by age, ASPECTS, stroke severity and time from onset to imaging. Clinicians and patients knew treatment assignment. Trained local assessors masked to allocation used structured functional interviews, while an independent imaging core laboratory reviewed baseline and follow-up scans. Independent medical monitors adjudicated serious adverse events.^{1,4}

The intervention allowed approved stent retrievers, aspiration catheters or a combination, chosen by the operator. Medical care followed guideline-based and local practice. Intravenous alteplase, when indicated, was to be given within three hours under the study's eligibility rules. One control received thrombectomy through team miscommunication. Two intervention patients did not undergo arterial puncture, one because an angiography suite was unavailable and one after spontaneous recanalisation. Four additional intervention patients underwent procedure initiation but no device thrombectomy, leaving 146 with a graded thrombectomy reperfusion result.^{1,4}

The primary efficacy outcome was mean utility-weighted modified Rankin scale at ninety days. Conventional Rankin scores zero through six were assigned weights of 10.0, 9.1, 7.6, 6.5, 3.3, zero and zero. Higher values indicate better function; severe dependence and death receive the same weight. The publication specified 0.3 points as a minimum clinically important difference. A Bayesian model adjusted for baseline ASPECTS

estimated the probability that thrombectomy improved the mean score. A posterior probability of at least 0.975 was required for success.¹

The maximum planned sample was 300, with simulated power of 90% for an assumed 1.25-point benefit. Interim analyses allowed stopping for futility or expected success, or restricting subsequent recruitment to ASPECTS four to five if the lower-score group appeared less promising. Neither stopping nor restriction occurred, and the trial completed its maximum sample. The primary analysis used Bayesian imputation based on baseline information and thirty-day function for missing ninety-day outcomes. One further control consent withdrawal left 299 patients in that analysis; 297 had observed ninety-day functional scores.^{1,4}

Secondary efficacy outcomes included the ordinal Rankin distribution, combining five and six; independence at Rankin zero to two; independent ambulation at zero to three; neurological improvement; quality of life; hospital stay; and follow-up imaging. Secondary comparisons used frequentist methods, with one-sided P values and an exploratory threshold below 0.025. No multiplicity correction was applied. Adverse events were analysed according to treatment received, defined by arterial puncture, creating groups of 151 and 149 rather than the randomised groups.¹

Median age was approximately 67 years. Median baseline stroke-severity scores were nineteen and eighteen, and about 65% had locally assessed ASPECTS four or five. Only 82 of 300 were randomised within six hours; 146 entered after twelve hours. Median onset-to-randomisation times were 10.9 and 12.6 hours. About 20% received alteplase. Diabetes was more common in the intervention group. Core-laboratory review placed 252 of 298 assessable scans within the intended ASPECTS range, with eleven below two and 35 above five.^{1,4}

Mean ninety-day utility-weighted scores were 2.93 with thrombectomy and 2.27 with medical care. The adjusted difference was 0.63 points, with a 95% credible interval from -0.09 to 1.34. The posterior probability of superiority was 0.96, below the required 0.975. Median conventional Rankin scores were five in both groups, and the secondary common odds ratio for better function was 1.40 (95% confidence interval 0.91–2.16).¹

Independence occurred in 22 of 151 intervention patients (14.6%) and thirteen of 146 controls (8.9%), an absolute difference of 5.7 percentage points (95% confidence interval -1.6 to 12.9). Independent ambulation occurred in 45 of 151 (29.8%) versus 29 of 146 (19.9%), difference 9.9 points (0.2–19.7). Major neurological improvement by day six occurred in 39 of 151 (25.8%) versus nineteen of 147 (12.9%). Median quality-of-life

ratings were sixty versus fifty on the reported zero-to-100 scale, while median hospital stay was eleven days in both groups.¹

Core-laboratory reperfusion of grade 2b or better was achieved in 107 of 146 patients (73.3%). Median follow-up infarct volumes were 166 versus 171 mL, without an appreciable between-group difference. In the as-treated safety analysis, ninety-day mortality was 53 of 150 (35.3%) versus 49 of 147 (33.3%), risk ratio 1.06 (0.77–1.45). Symptomatic intracranial haemorrhage at 24 hours occurred in six of 151 (4.0%) versus two of 149 (1.3%), risk ratio 2.96 (0.61–14.43). Parenchymal haematoma and subarachnoid haemorrhage were more frequent after intervention.¹

Decompressive hemicraniectomy occurred in 33 of 151 intervention recipients (21.9%) and 22 of 149 controls (14.8%), with an imprecise difference. Nine patients had a procedural complication, including embolisation into a new territory, arterial perforation or dissection; some had more than one complication. One hundred patients in each group had at least one serious adverse event. Sensitivity analysis restricted to core-laboratory ASPECTS two to five produced a larger utility-weighted difference, 0.82 points (95% credible interval 0.03–1.60), with posterior probability 0.980.^{1,4}

Critique

The main conclusion is that TESLA did not meet its criterion for primary efficacy. That conclusion should not be converted into evidence that thrombectomy has no benefit in this setting. The point estimate was more than twice the report's minimum clinically important difference, and the credible interval included both negligible benefit and a substantial improvement. Conversely, a 96% posterior probability does not justify retrospectively lowering the 97.5% threshold. Reporting the estimate, interval and failed success criterion together is the most informative interpretation.¹

The assumed treatment effect helps explain the remaining uncertainty. TESLA was designed around a 1.25-point improvement, approximately twice the adjusted effect actually observed. It completed its intended maximum sample and was not stopped prematurely. The problem is therefore limited precision for a smaller effect than the one used in planning. The editorial considers an optimistic effect assumption a plausible reason for a false-negative result. This is a reasonable explanation, but neither post hoc power language nor consistency with other studies can establish that the true effect must be positive.^{1,3}

The utility-weighted endpoint makes explicit that changes between Rankin categories do not have equal value. It can capture improvements missed by a simple independence

threshold. However, its mean is not a patient's actual Rankin score, and a 0.63-point difference cannot be translated into a uniform degree of recovery for every recipient. Giving severe dependence and death identical weights also requires mortality and the full distribution to be reported separately. The published 0.3-point clinical-importance threshold should be understood on this transformed scale.¹

The secondary results are broadly supportive but not decisive. More intervention patients reached independent ambulation and major early neurological improvement. These outcomes were among numerous unadjusted comparisons, and the reported P values were one-sided. A conventional two-sided P-value convention cannot simply be imposed on the table without recognising the stated testing rules. In particular, a favourable secondary interval cannot overturn the primary result. The patient-facing message remains that a meaningful functional advantage is plausible, with its size uncertain.¹

Randomisation, masked functional assessment and central imaging review are major strengths. Follow-up was nearly complete, reducing concern that missing outcomes drove the primary estimate. The treatment could not be concealed from clinical teams, however, and postprocedural decisions may have differed. More intervention recipients underwent hemicraniectomy. This could reflect greater reperfusion-related swelling, more aggressive rescue care or both; the trial does not identify the mechanism. Such co-interventions are relevant when interpreting the whole treatment pathway rather than device recanalisation in isolation.¹

TESLA's imaging strategy is both its principal contribution and a source of uncertainty. Local non-contrast CT interpretation closely resembles real clinical selection, but central reviewers placed approximately one-sixth of assessable patients outside the intended ASPECTS range. Restriction to the centrally confirmed range produced a posterior probability exceeding the primary threshold. This supports further interest in the target population, but remains a sensitivity analysis with fewer patients. It cannot replace the main analysis of the locally selected strategy that the trial set out to test.^{1,4}

Nor does the difference between local and central scoring demonstrate that automated imaging would solve selection. Investigators were specifically prohibited from using automated ASPECTS to enrol patients. The supplementary automated-score work was largely retrospective, with only moderate agreement between methods, and the planned prospective software-defined efficacy analysis was not performed. TESLA therefore provides no randomised comparison of manual and automated selection, or of basic imaging against a mandatory perfusion pathway.⁴

Timing is a plausible explanation for differences from other trials. Almost three-quarters were randomised beyond six hours, and nearly half beyond twelve. The early subgroup had larger apparent gains in independence than the later subgroup. However, those binary early-versus-late comparisons were post hoc, and the primary-outcome subgroup analysis did not establish a treatment interaction. A significant result in one subgroup and a non-significant result in another would not, by itself, prove that their treatment effects differ. TESLA leaves late-window selection uncertain rather than establishing a sharp six-hour boundary.^{1,4}

Procedural performance may also matter. Successful core-laboratory reperfusion was achieved in 73%, and the median interval from imaging to puncture was approximately ninety minutes. These are features of the care actually delivered. They could influence the observed average effect, but comparisons against faster or more successful programmes are not randomised comparisons. Likewise, the statistically significant variation in treatment effects between sites deserves attention without being attributed automatically to operator quality. Small site samples, case mix and chance can contribute to apparent heterogeneity.^{1,4}

Safety cannot be summarised as proven equivalence. Mortality was similar in observed proportions, but its confidence interval allowed clinically important benefit or harm. Symptomatic haemorrhage was uncommon and numerically more frequent with intervention, with very wide uncertainty. More parenchymal and subarachnoid bleeding was established in exploratory radiographic comparisons. These findings do not negate a possible functional benefit, but they matter when the average benefit is modest and the probability of independence remains low.¹

The as-treated safety population also deserves explicit labelling. Reassigning the few crossovers by arterial puncture gives a clinically useful account of procedural exposure, but no longer fully preserves randomisation. It explains why safety denominators differ from those in the functional-outcome table and is not itself a reporting error. Similarly, the nine patients with complications should not be confused with the sum of individual complication types, because several patients experienced more than one.^{1,4}

Generalisability is broader geographically and institutionally than a highly selected single-centre series, but remains bounded by trial eligibility. Participants were previously independent, no older than 85 and free of the specified severe swelling and access limitations at selection. Screening logs were not required, so the proportion of otherwise eligible patients treated outside the study cannot be quantified. The study received unrestricted industry funding, with no reported funder role in design, conduct, analysis or publication; independent statistical and clinical oversight are relevant safeguards.¹

Body of Evidence

RESCUE-Japan LIMIT first provided randomised evidence that a large imaging-defined infarct need not make thrombectomy futile. Its population had ASPECTS three to five and was largely selected using MRI. It supplied a reason to test broader populations, while not answering TESLA's specific question about a non-contrast CT pathway through 24 hours.^{1,5}

SELECT2 and ANGEL-ASPECT then demonstrated better functional outcomes using combinations of ASPECTS and volume-based criteria. Their later treatment windows are relevant to TESLA, but their imaging eligibility differed. An MRI- or perfusion-defined core does not guarantee equally extensive visible change on non-contrast CT. Thus, consistency in the direction of benefit across these studies is informative without making their patients biologically identical.^{1,6,7}

TENSION is the closest earlier comparison because it predominantly used non-contrast CT and demonstrated an ordinal functional benefit in 253 patients. Its treatment window was shorter, and a much greater proportion entered early. TESLA's more frequent late presentation is therefore a credible source of difference between results. It remains a hypothesis derived from comparisons between trials, rather than proof that timing alone explains the smaller TESLA estimate.^{1,2}

LASTE, published in 2024, demonstrated a clear functional benefit in an early-presenting population with very extensive injury, predominantly assessed by MRI. Its generalised odds ratio was 1.63 (95% confidence interval 1.29–2.06), with lower mortality. LASTE included many patients with ASPECTS two or lower, while retaining restrictions concerning age, prior function and mass effect. It strengthens the case against treating large infarct size alone as futility, but does not resolve TESLA's question about later, non-contrast CT-defined injury.⁸

The two 2024 trials should therefore be read together without forcing them into simple opposition. LASTE provides strong evidence for treatment in its selected population. TESLA provides a less precise estimate for a different pathway, with a primary interval overlapping clinically relevant benefit. Calling one positive and the other negative captures their formal thresholds but loses information about effect size, uncertainty and patient selection. Their difference cannot identify which imaging modality is superior because patients were not randomised between imaging strategies.^{1,8}

The accompanying editorial places TESLA within this generally favourable large-infarct evidence base and emphasises the potential access benefits of simpler imaging. It also refers to one-year findings presented at a conference. Those findings had a different

evidential status from the published ninety-day primary report and should not be used to relabel that analysis as positive. The possibility of recovery continuing beyond ninety days is clinically relevant, but a later endpoint answers an additional question rather than erasing the original one.^{1,3}

For an individual patient, the remaining uncertainty has practical consequences. Extensive infarction predicts a high likelihood of death or dependence even when intervention is offered. A potential gain in walking or self-care may still be valuable, but its value depends on the person's preferences and the burdens they would accept. Neither a low probability of independence nor a non-significant trial result automatically settles that judgement. Discussions should use absolute outcomes alongside the evidence supporting treatment in the closest available population.^{1,3}

By the end of 2024, the combined evidence supported broadening thrombectomy assessment for large-infarct stroke. TESLA added an important qualification: its pragmatic non-contrast CT selection strategy across 24 hours did not independently demonstrate the planned functional benefit. It should encourage careful attention to timing, image interpretation and procedural delivery, while preserving uncertainty about the later window. Further clarification of these boundaries is more useful than either universal exclusion of large infarcts or an assumption that all such patients benefit equally.^{1,3,8}

Does thrombectomy improve recovery when a large infarct is selected by non-contrast CT within 24 hours?

TESLA did not meet its primary success threshold, but its estimate remained compatible with a clinically important benefit. Uncertainty is greatest when applying the result across the full late treatment window.

References

1. Yoo AJ, Zaidat OO, Sheth SA, Rai AT, Ortega-Gutierrez S, Given CA II, et al.; Writing Committee for the TESLA Investigators. Thrombectomy for Stroke With Large Infarct on Noncontrast CT: The TESLA Randomized Clinical Trial. *JAMA* 2024;332(16):1355–1366.
2. Bendszus M, Fiehler J, Subtil F, Bonekamp S, Aamodt AH, Fuentes B, et al.; TENSION Investigators. Endovascular thrombectomy for acute ischaemic stroke with established large infarct: multicentre, open-label, randomised trial. *Lancet* 2023;402:1753–1763.

3. Khalife J, Tonetti DA, Jovin TG. CT-Guided Thrombectomy for Large Core Stroke Up to 24 Hours—Another Piece in a Complex Puzzle. *JAMA* 2024;332(16):1338–1339.
4. TESLA Investigators. Supplementary online content to: Thrombectomy for Stroke With Large Infarct on Noncontrast CT: The TESLA Randomized Clinical Trial. *JAMA* 2024. doi:10.1001/jama.2024.13933.
5. Yoshimura S, Sakai N, Yamagami H, Uchida K, Beppu M, Toyoda K, et al.; RESCUE-Japan LIMIT Investigators. Endovascular Therapy for Acute Stroke with a Large Ischemic Region. *N Engl J Med* 2022;386(14):1303–1313.
6. Sarraj A, Hassan AE, Abraham MG, Ortega-Gutierrez S, Kasner SE, Hussain MS, et al.; SELECT2 Investigators. Trial of Endovascular Thrombectomy for Large Ischemic Strokes. *N Engl J Med* 2023;388(14):1259–1271.
7. Huo X, Ma G, Tong X, Zhang X, Pan Y, Nguyen TN, et al.; ANGEL-ASPECT Investigators. Trial of Endovascular Therapy for Acute Ischemic Stroke with Large Infarct. *N Engl J Med* 2023;388(14):1272–1283.
8. Costalat V, Jovin TG, Albucher JF, Cognard C, Henon H, Nouri N, et al.; LASTE Trial Investigators. Trial of Thrombectomy for Stroke with a Large Infarct of Unrestricted Size. *N Engl J Med* 2024;390(18):1677–1689.

INTERACT4

Li G, Lin Y, Yang J, Anderson CS, Chen C, Liu F, et al.; INTERACT4 Investigators. Intensive Ambulance-Delivered Blood-Pressure Reduction in Hyperacute Stroke. *N Engl J Med* 2024;390(20):1862–1872.

Introduction

An ambulance offers an opportunity to treat acute stroke during its earliest and potentially most modifiable phase. It also presents a fundamental diagnostic limitation: without brain imaging, clinicians cannot reliably distinguish intracerebral haemorrhage from cerebral ischaemia. Lowering an elevated blood pressure could limit continued bleeding in the former while reducing perfusion of threatened brain in the latter. A treatment directed at the same physiological measurement may therefore have different consequences across the patients who share a prehospital diagnosis of suspected stroke.¹

Hospital trials had provided reasons to investigate early blood-pressure reduction, particularly in intracerebral haemorrhage, but had not established that progressively lower pressures are uniformly beneficial. INTERACT2 suggested a modest improvement across the disability scale despite a neutral primary binary endpoint. ATACH-2 found no functional benefit from a more intensive target and identified more renal adverse events. These findings concerned patients whose haemorrhage had already been diagnosed, leaving uncertainty about treatment before imaging and substantially earlier after symptom onset.^{2,3}

Previous ambulance trials had tested transdermal glyceryl trinitrate. RIGHT-2 did not improve overall functional outcome, and MR ASAP stopped early following concern about harm in intracerebral haemorrhage. Their results challenged the proposition that a readily delivered antihypertensive could benefit undifferentiated stroke. They did not directly answer whether rapid intravenous treatment with a different agent and a defined pressure target would behave similarly.^{4,5}

INTERACT4 tested that strategy in Chinese ambulance services, initiating treatment within two hours of suspected stroke. Its clinical question was deliberately broader than whether blood-pressure reduction helps a known haemorrhage: could treating hypertension before diagnostic clarification improve ninety-day function across the whole suspected-stroke population? The answer needed to account for the overall

randomised comparison and the possibility that an apparently neutral average concealed benefit and harm in different underlying diseases.¹

Synopsis

INTERACT4 was an open-label, randomised superiority trial with masked outcome assessment, conducted through 51 hospitals in China. Forty-three used hospital-based ambulances and eight centralised services; the ambulances were staffed by physicians. Recruitment ran from 20 March 2020 to 31 August 2023. Academic and public funding supported the study, with additional sponsorship from Takeda Pharmaceuticals China, manufacturer of urapidil. The report states that the company had no role in trial conduct. An independent data and safety monitoring board oversaw the trial.¹

Adults were eligible with suspected acute stroke, a face-arm-speech-time score of at least two including an arm motor deficit, systolic pressure of at least 150 mmHg on two measurements, and the ability to start treatment within two hours of symptom onset or last known well. Important exclusions included unresponsive coma, severe comorbidity or dependence in daily activities before stroke, epilepsy or seizure at onset, recent head injury and marked hypoglycaemia. Imaging confirmation was not required before randomisation.^{1,6}

A secure web system allocated treatment equally, using minimisation for geographic region, age below or at least 65 years, and baseline face-arm-speech-time score. At fifty sites, a consent waiver allowed immediate treatment; one used brief written consent. Written patient or surrogate consent for data use and follow-up followed in hospital. Ambulance and hospital clinicians knew allocation. Trained assessors unaware of treatment assignment assessed ninety-day function by telephone or in person, often using information from proxies.^{1,6}

The intervention targeted systolic pressure of 130–140 mmHg within thirty minutes, maintained until hospital arrival. Intravenous urapidil, 25 mg over one minute, could be repeated once after five minutes if pressure remained high. Patients were kept horizontal and pressure was monitored at five-minute intervals; treatment stopped at the lower threshold. Usual-care patients received ambulance antihypertensives only for systolic pressure at least 220 mmHg or diastolic pressure at least 110 mmHg. After arrival, both groups received hospital care according to stroke diagnosis and usual management.^{1,6}

The primary endpoint was the seven-level modified Rankin scale at ninety days, ranging from no symptoms to death. A mixed ordinal logistic model adjusted for the allocation

factors and included a site random effect. Its common odds ratio expressed the odds of worse function, so values below one favoured intervention. Secondary outcomes included death or dependency, mortality, neurological impairment, quality of life, early discharge and institutional residence. Serious adverse events were assessed separately. Reported subgroup confidence intervals were not adjusted for multiplicity.^{1,6}

The published supplement describes a blinded sample-size reassessment after fewer stroke mimics than anticipated were observed. The revised target of 2,320 incorporated a 6% mimic assumption, rather than 30%, and allowed 90% power for a common odds ratio of 0.78 with 5% missing outcomes. This was a reassessment of expected diagnostic composition, rather than stopping because efficacy had emerged. The published analysis used observed primary outcomes because missingness remained below its stated 5% imputation threshold.^{1,6}

Of 5,488 screened patients, 2,425 were randomised, 1,215 to intervention and 1,210 to usual care. Consent was not obtained for twenty patients and one intervention participant withdrew consent, leaving an intention-to-treat population of 2,404: 1,205 and 1,199. Ninety-day Rankin scores were available for 1,185 and 1,177. Forty-one patients were lost to follow-up; one further control declined functional assessment but was known to be alive, explaining the additional patient in the mortality denominator.^{1,6}

Mean age was seventy years, 61.7% were men and approximately 99% were Han Chinese. Median time to randomisation was 61 minutes and mean baseline pressure was 178/98 mmHg. Final diagnoses included cerebral ischaemia in 1,199 patients, encompassing 53 transient ischaemic attacks, and haemorrhagic stroke in 1,041, including twelve subarachnoid haemorrhages. There were 154 mimics and ten patients without a final diagnosis. Of 1,029 intracerebral haemorrhages, 84.7% were deep; median initial volumes were seventeen and twenty millilitres in the two groups.^{1,6}

Ambulance antihypertensive treatment was given to 1,078 intervention patients (89.5%) and 116 controls (9.7%). Urapidil accounted for 98.2% of treated patients. Mean systolic pressure on hospital arrival was 159 versus 170 mmHg; the modelled prehospital difference was -15 mmHg, with a much smaller difference after admission. At 24 hours, mean systolic pressure was 140 mmHg in both groups. Hospital intravenous antihypertensive use was somewhat greater after intervention. The appendix records intravenous thrombolysis in 274 versus 256 patients and thrombectomy in 75 versus 61.^{1,6}

The primary comparison was neutral: common odds ratio for worse ninety-day function 1.00 (95% confidence interval 0.87–1.15). Death or dependency, defined as Rankin three

to six, occurred in 702 of 1,185 intervention patients (59.2%) and 721 of 1,177 controls (61.3%), adjusted odds ratio 0.92 (0.78–1.09). Mortality was 267 of 1,185 (22.5%) versus 266 of 1,178 (22.6%), adjusted odds ratio 1.00 (0.82–1.22). Neurological impairment, quality of life, early discharge and institutional residence showed no persuasive overall improvement.^{1,6}

The reported stroke-type subgroup estimates pointed in opposite directions. For haemorrhagic stroke, the common odds ratio for worse function was 0.75 (0.60–0.92); for cerebral ischaemia it was 1.30 (1.06–1.60). Observed mortality within these groups was respectively 125 of 515 versus 156 of 507, and 133 of 592 versus 101 of 592. These subtype findings were outside a hierarchical confirmatory testing strategy. The authors explicitly cautioned against drawing causal inferences from the subgroup associations.^{1,6}

There were 350 versus 351 serious adverse events, affecting 331 intervention patients (27.5%) and 344 controls (28.7%). In haemorrhagic stroke, fewer patients had a serious adverse event after intervention; in cerebral ischaemia, more did. A per-protocol analysis, excluding 421 patients for reported violations, remained neutral overall, as did analyses excluding mimics and selected alternative diagnoses. In the subset with assessable serial haemorrhage imaging, absolute haematoma growth of at least 6 mL occurred in 42 of 371 versus 66 of 362, an exploratory finding consistent with the haemorrhagic subgroup signal.⁶

Critique

INTERACT4 addressed a real decision at the point it arises. Randomisation before imaging tested the policy of treating suspected stroke in an ambulance, rather than a retrospectively selected group whose diagnosis was already known. Its large sample, multiple sites, central allocation and masked functional assessment strengthen the primary comparison. Recruitment approximately one hour after onset also created a substantial opportunity to influence early disease processes. For an intervention intended to act before hospital admission, feasibility at that time is a meaningful achievement.^{1,6}

The overall result gives a clear answer to that policy question. There was no demonstrated improvement in ninety-day functional outcome from routinely initiating this intensive strategy in the enrolled suspected-stroke population. The confidence interval excludes the large average benefit used in the sample-size calculation, although it remains compatible with smaller benefit or harm. It should not be translated into proof of equivalence. Nor should the P value of 0.16 printed beside the primary results be

interpreted as the treatment-effect P value: it describes the test of the proportional-odds assumption.¹

The most important qualification is biological heterogeneity. Intracerebral bleeding and arterial occlusion are sufficiently different that opposing responses to blood-pressure reduction are credible. The haemorrhagic and ischaemic functional distributions, their mortality patterns and their serious-adverse-event directions are mutually supportive observations. Yet these are correlated outcomes in the same patients, not independent replications. The subgroup intervals were unadjusted for multiplicity, and the neutral primary comparison cannot be replaced with two definitive positive trials selected from within it.^{1,6}

These qualifications call for calibrated interpretation rather than dismissal of the subgroup findings. A plausible signal of harm matters when considering implementation of a treatment with no overall proven benefit. Conversely, the haemorrhagic signal makes diagnosis-specific ultra-early treatment a compelling research question. INTERACT4 does not establish a clinical rule that can reliably identify such patients before imaging. A policy requiring prehospital CT confirmation would involve a different diagnostic pathway and a differently selected population from the policy randomised here.^{1,7}

The achieved physiological contrast also needs careful description. The target was 130–140 mmHg, but mean arrival systolic pressure remained 159 mmHg. Thus, the trial assessed an attempt at rapid intensive reduction, not consistent attainment of the target before hospital. Most intervention patients received treatment and there was a substantial early between-group separation, making wholesale failure to deliver the intervention unlikely. Nonetheless, neither the benefits nor the harms of a more reliably attained or more prolonged low target can be inferred from these results.^{1,6}

The experiment did not isolate the effect of one ambulance dose from everything that followed. Hospital staff knew allocation, post-admission intravenous antihypertensive use differed, and a smaller blood-pressure separation persisted early after arrival. These differences form part of the practical effect of assigning the treatment strategy, but complicate attributing every outcome exclusively to prehospital exposure. Subsequent thrombolysis, thrombectomy, supportive care and treatment limitations could also influence recovery. The reported management data help assess these possibilities without proving that all cointerventions were identical.⁶

Missing functional outcomes were few and reasonably balanced, which limits attrition bias. The exclusion of twenty-one randomised patients for consent reasons prevents a

completely inclusive intention-to-treat analysis, but represents a small fraction of allocation. The per-protocol analysis is less secure as an estimate of efficacy under perfect adherence: exclusions after randomisation can select patients according to clinical course or operational circumstances. Its neutral result supports the main finding, but does not restore the protection of randomisation within the selected adherent population.^{1,6}

The imaging findings provide mechanistic support with important limits. Follow-up haematoma-growth data were available in 733 patients, fewer than the 1,029 with intracerebral haemorrhage. Availability of serial scans may depend on survival, deterioration or clinical decisions. The current report used investigator-reported imaging abnormalities, and initial scans were obtained after the intervention had begun. Consequently, lower subsequent expansion is compatible with an anti-bleeding effect but is not a complete, independently established explanation for the functional subgroup difference.^{1,6}

The accompanying editorial questioned the imaging characterisation of the ischaemic population and the generalisability of the prehospital findings. That concern deserves attention without turning an uncertain table interpretation into an accusation of erroneous timing. The supplementary categories do not clearly establish that every reported infarct appearance was mutually exclusive or obtained using an identical modality and time point. A more detailed imaging account would improve clinical interpretation, but the available material does not justify reconstructing unreported onset times or concluding that the enrolled population was misclassified.^{6,7}

External validity depends on more than nationality. The cohort had a high proportion of haemorrhagic stroke, predominantly deep haemorrhage, and was treated by physician-staffed ambulances in a system using urapidil extensively. The distribution of diagnoses, ambulance staffing, transit intervals and subsequent stroke services may differ elsewhere. The overall estimate could change if an intervention truly benefits one diagnosis and harms another, even without any change in its diagnosis-specific effects. This is a reason to examine local case mix, not to calculate a new net benefit from unconfirmed subgroup effects.^{1,6}

Finally, similar total serious-adverse-event rates do not establish safety for every stroke subtype. Functional worsening or death can arise from the underlying stroke without being recognised as a drug reaction, and an aggregate safety total can conceal opposing patterns. The trial supports restraint about routine intensive reduction before diagnostic clarification, while preserving a potentially important hypothesis for patients with early

confirmed haemorrhage. It does not resolve the optimal agent, rate of reduction or target for all acute stroke presentations.^{1,6}

Body of Evidence

INTERACT2 provided much of the rationale for pursuing earlier treatment in intracerebral haemorrhage. It randomised 2,839 patients within six hours to systolic targets below 140 or below 180 mmHg. Death or major disability occurred in 52.0% versus 55.6%, odds ratio 0.87 (0.75–1.01), with $P=0.06$. A prespecified ordinal analysis favoured intensive treatment modestly. This supported further study without proving that the same strategy would help an unselected ambulance population.²

ATACH-2 tested lower achieved pressures in 1,000 patients with confirmed intracerebral haemorrhage, targeting 110–139 versus 140–179 mmHg within 4.5 hours. Death or disability was similar, with adjusted relative risk 1.04 (0.85–1.27), while renal adverse events were more frequent after intensive treatment. Together, these trials caution against treating blood-pressure reduction as an intervention whose benefit simply increases with intensity. Their targets, timing and diagnostic certainty differed from INTERACT4.³

INTERACT3 subsequently tested a hospital care bundle incorporating blood-pressure control, glucose and temperature management, and correction of abnormal anticoagulation. Its stepped-wedge cluster trial included 7,036 patients with intracerebral haemorrhage and found better ordinal functional outcome with the bundle, common odds ratio for worse outcome 0.86 (0.76–0.97). This supported organised early haemorrhage care. Because several components were implemented together, it did not identify the independent effect of blood-pressure lowering or establish that treatment should precede imaging.⁸

RIGHT-2 was closer to INTERACT4 operationally. UK paramedics randomised 1,149 patients with presumed stroke within four hours to a glyceryl trinitrate patch or sham treatment. It achieved a modest pressure reduction without improving ninety-day function overall; the common odds ratio for worse outcome in all randomised patients was 1.04 (0.84–1.29). Its lower haemorrhage proportion, more frequent mimics, transdermal agent and smaller physiological separation limit a direct comparison with intravenous urapidil.⁴

MR ASAP tested prehospital glyceryl trinitrate within three hours in the Netherlands. Recruitment stopped after 380 randomisations because of concern about early mortality in intracerebral haemorrhage; 325 patients entered its principal analysis. Overall

function was neutral, common odds ratio 0.97 (0.65–1.47). The early haemorrhage safety signal was imprecise and did not translate into a clear ninety-day mortality excess. Nevertheless, its direction differed from INTERACT4's haemorrhagic finding, arguing against assuming that all antihypertensive agents or delivery strategies have interchangeable effects.⁵

Evidence from ischaemic stroke also discouraged a universal low target. OPTIMAL-BP randomised patients after successful thrombectomy to systolic pressure below 140 or 140–180 mmHg. Functional independence was less frequent with intensive management, 39.4% versus 54.4%, adjusted odds ratio 0.56 (0.33–0.96). This was a post-reperfusion hospital trial, so it cannot directly establish the mechanism or magnitude of INTERACT4's ischaemic signal. It does, however, reinforce the need to test a pressure target within its particular disease stage.⁹

By the end of 2024, the evidence therefore supported a more specific interpretation of “early treatment”. Prompt, organised care for confirmed haemorrhage had evidence behind it, while greater intensity and treatment before diagnosis were separate propositions. INTERACT4 supplied the largest direct test of intensive intravenous pressure reduction in suspected stroke before imaging and did not demonstrate overall benefit. Its opposing subtype findings sharpened the next question: whether very early treatment selected by a reliable haemorrhage diagnosis can improve outcomes without exposing patients with ischaemia to avoidable harm.¹

Should intensive blood-pressure reduction begin routinely in the ambulance before stroke type is known?

No overall functional benefit was demonstrated. The opposing exploratory signals in haemorrhagic and ischaemic stroke support diagnosis-specific investigation rather than routine treatment of undifferentiated stroke.

References

1. Li G, Lin Y, Yang J, Anderson CS, Chen C, Liu F, et al.; INTERACT4 Investigators. Intensive Ambulance-Delivered Blood-Pressure Reduction in Hyperacute Stroke. *N Engl J Med* 2024;390(20):1862–1872.
2. Anderson CS, Heeley E, Huang Y, Wang J, Stapf C, Delcourt C, et al.; INTERACT2 Investigators. Rapid Blood-Pressure Lowering in Patients with Acute Intracerebral Hemorrhage. *N Engl J Med* 2013;368(25):2355–2365.

3. Qureshi AI, Palesch YY, Barsan WG, Hanley DF, Hsu CY, Martin RL, et al.; ATACH-2 Trial Investigators and the Neurological Emergency Treatment Trials Network. Intensive Blood-Pressure Lowering in Patients with Acute Cerebral Hemorrhage. *N Engl J Med* 2016;375(11):1033–1043.
4. RIGHT-2 Investigators. Prehospital transdermal glyceryl trinitrate in patients with ultra-acute presumed stroke (RIGHT-2): an ambulance-based, randomised, sham-controlled, blinded, phase 3 trial. *Lancet* 2019;393:1009–1020.
5. van den Berg SA, Uniken Venema SM, Reinink H, Hofmeijer J, Schonewille WJ, Miedema I, et al.; MR ASAP Investigators. Prehospital transdermal glyceryl trinitrate in patients with presumed acute stroke (MR ASAP): an ambulance-based, multicentre, randomised, open-label, blinded endpoint, phase 3 trial. *Lancet Neurol* 2022;21(11):971–981.
6. INTERACT4 Investigators. Supplementary appendix to: Intensive Ambulance-Delivered Blood-Pressure Reduction in Hyperacute Stroke. *N Engl J Med* 2024. doi:10.1056/NEJMoa2314741.
7. Edlow JA. Lowering Blood Pressure in Stroke Patients in the Ambulance—A Bridge Too Close? *N Engl J Med* 2024;390(20):1916–1917.
8. Ma L, Hu X, Song L, Chen X, Ouyang M, Billot L, et al.; INTERACT3 Investigators. 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.
9. Nam HS, Kim YD, Heo J, Lee H, Jung JW, Choi JK, et al.; OPTIMAL-BP Trial Investigators. Intensive vs Conventional Blood Pressure Lowering After Endovascular Thrombectomy in Acute Ischemic Stroke: The OPTIMAL-BP Randomized Clinical Trial. *JAMA* 2023;330(9):832–842.

Possible Reporting Inconsistencies. Supplementary Table S19 lists control-group mortality in cerebral ischaemia as 101/474 (17.0%), whereas its Rankin distribution and Table S20 report 101/592 (17.1%). There appears to be a denominator error in the mortality row; this chapter uses the consistently reported Rankin counts.

TIMELESS

Albers GW, Juma M, Purdon B, Zaidi SF, Streib C, Shuaib A, et al.; TIMELESS Investigators. Tenecteplase for Stroke at 4.5 to 24 Hours with Perfusion-Imaging Selection. N Engl J Med 2024;390(8):701–711.

Introduction

An occluded cerebral artery does not produce irreversible injury at the same rate in every patient. Some people retain substantial potentially salvageable tissue many hours after they were last known to be well. Perfusion imaging can identify a relatively small estimated infarct core surrounded by a larger region of hypoperfusion, providing a biological rationale for reperfusion beyond conventional thrombolysis windows. The treatment question then depends on which reperfusion therapy is available and what additional benefit a drug can offer alongside it.¹

Trials of imaging-selected alteplase had established that an uncertain or late presentation did not invariably preclude benefit. WAKE-UP used a magnetic resonance imaging pattern suggesting recent infarction, while EXTEND used perfusion mismatch to select patients for treatment beyond 4.5 hours or after waking with stroke. These strategies selected different populations and did not establish that every thrombolytic regimen could be extended to twenty-four hours.^{2,3}

Tenecteplase is administered as a single bolus, making it attractive when rapid treatment and interhospital transfer are priorities. EXTEND-IA TNK found more early reperfusion with tenecteplase than alteplase before thrombectomy within the conventional early window. Whether this advantage would translate into better recovery when treatment started much later remained uncertain. A drug might open an artery before catheter treatment, improve distal perfusion or offer an alternative when thrombectomy was unavailable. Conversely, rapid successful thrombectomy might leave little additional benefit for preceding thrombolysis.^{1,4}

TIMELESS tested tenecteplase against placebo between 4.5 and twenty-four hours after last known well in patients with an anterior-circulation large-vessel occlusion and favourable perfusion imaging. Most also underwent thrombectomy. Its interpretation therefore turns on the distinction between extending intravenous thrombolysis as a primary reperfusion treatment and adding it shortly before a highly effective endovascular procedure.¹

Synopsis

TIMELESS was a phase 3, double-blind, placebo-controlled randomised trial conducted through 112 centres in the United States and Canada. Recruitment ran from March 2019 to December 2022. Genentech funded the trial, participated in its design and data oversight, supplied treatment and undertook analyses through two employee authors. An independent monitoring committee supervised safety and efficacy; an independent statistical centre performed the interim analysis. The academic first author had unrestricted access to the final data, and the sponsor could not prevent or delay publication.^{1,5}

Eligible adults had an internal carotid or M1 or M2 middle cerebral artery occlusion, a National Institutes of Health Stroke Scale score of at least five, and a prestroke modified Rankin score of zero to two. Treatment had to start 4.5–24 hours after last known well, including unwitnessed and wake-up strokes. Computed tomography or magnetic resonance perfusion imaging, processed with automated software, had to show an estimated core below seventy millilitres, a hypoperfused-tissue-to-core ratio of at least 1.8, and an absolute mismatch of at least fifteen millilitres. Qualifying imaging generally had to precede treatment by no more than ninety minutes.^{1,5}

A central web system assigned patients equally to intravenous tenecteplase, 0.25 mg/kg to a maximum of 25 mg, or a matching placebo bolus over approximately five seconds. Randomisation used permuted blocks and stratification for age, occlusion site, stroke severity and whether the site could perform thrombectomy. Patients, clinical teams and outcome assessors were masked. Thrombectomy remained part of usual care and was anticipated for eligible internal carotid and M1 occlusions; its use for M2 occlusions was discretionary. The study neither compared tenecteplase with alteplase nor required clinicians to withhold thrombectomy.^{1,5}

The primary endpoint was the seven-category modified Rankin scale at ninety days, from no symptoms to death. The adjusted common odds ratio described a shift towards better outcomes, with values above one favouring tenecteplase. Secondary endpoints included functional independence, arterial recanalisation, tissue reperfusion and other measures of recovery. A sequential testing scheme stopped formal testing of secondary efficacy outcomes if the primary comparison was not significant. Their confidence intervals therefore describe exploratory estimates without protection against multiple testing.^{1,5}

The planned sample of 456 provided 90% power for an assumed common odds ratio of 1.76. One interim analysis used a conservative stopping boundary, leaving a final significance threshold of 0.049. The published primary model adjusted for age, occlusion site and protocol version. At the journal's request, multiple imputation using one

hundred datasets replaced last-observation-carried-forward imputation for the principal analysis. The article's displayed Rankin counts retained the latter approach; three patients without any postbaseline assessment were excluded from efficacy analysis.^{1,5}

The trial randomised 458 patients, 228 to tenecteplase and 230 to placebo, and completed its planned recruitment. The primary analysis included 455, with eighteen missing ninety-day Rankin scores imputed. Median age was seventy-two and seventy-three years, 53.5% were women, and baseline median stroke severity was twelve in both groups. Median estimated core volumes were seven and five millilitres. Median times from last known well to treatment were 12.7 and 13.0 hours, indicating a selected population with relatively little established infarction despite long elapsed time.^{1,5}

Thrombectomy was planned in 400 patients and the clot was accessed in 354, or 77.3% of all participants. Only sixteen patients were randomised at centres without thrombectomy capability. Median time from study drug to arterial puncture was fifteen minutes with tenecteplase and seventeen with placebo; the corresponding times to clot contact were thirty-three and 38.5 minutes. Of the patients for whom thrombectomy was planned but not completed, some improved or became unsuitable, some had no visible clot at angiography and others had a clot that could not be reached. These circumstances cannot all be classified as successful pharmacological recanalisation.^{1,5}

Median ninety-day Rankin score was three in each group. The primary adjusted common odds ratio was 1.13 (95% confidence interval 0.82–1.57; $P=0.45$), providing no statistically significant evidence of improved functional outcome. Functional independence, a Rankin score of zero to two, was displayed in 104 of 226 patients given tenecteplase and 97 of 229 given placebo, or 46.0% and 42.4%; the adjusted odds ratio was 1.18 (0.80–1.74). The per-protocol analysis was also neutral.^{1,5}

Complete arterial recanalisation at twenty-four hours was observed in 148 of 193 assessed patients given tenecteplase and 124 of 194 given placebo, or 76.7% and 63.9%; the adjusted odds ratio was 1.89 (1.21–2.95). However, substantial tissue reperfusion at twenty-four hours occurred in 56.9% and 57.7%, respectively, with an adjusted odds ratio of 1.04 (0.69–1.57). Successful procedural reperfusion occurred in 89.1% and 85.4% of those assessed after thrombectomy. These exploratory imaging findings did not establish a corresponding clinical benefit.^{1,5}

The M1 occlusion subgroup showed a common odds ratio of 1.59 (1.00–2.52), and its exploratory independence comparison also favoured tenecteplase. Other occlusion subgroups did not show the same pattern. Neither these estimates nor the imaging

findings were adjusted for multiplicity. Among patients who actually did not undergo thrombectomy, the common odds ratio was 1.04 (0.52–2.06), a highly imprecise comparison within a subgroup defined partly by postrandomisation events.^{1,5}

Twenty-six patients did not receive study medication, and one placebo-assigned patient inadvertently received tenecteplase. Safety analyses therefore used treatment received in 218 tenecteplase and 214 placebo recipients. Ninety-day mortality was 43/218, or 19.7%, versus 39/214, or 18.2%. Symptomatic intracranial haemorrhage within thirty-six hours occurred in seven patients, or 3.2%, versus five, or 2.3%. Serious adverse events affected 41.7% and 47.7%. These small haemorrhage counts and descriptive safety comparisons did not demonstrate equivalent safety.^{1,5}

Critique

TIMELESS directly addressed whether late tenecteplase improves patient-centred recovery in a carefully selected population. Central allocation, matching placebo, masking and independent imaging assessment reduced several important sources of bias. The Rankin scale captured the whole range from recovery to death, and the trial reached its planned sample rather than stopping after an unstable early efficacy signal. These features give the neutral primary comparison considerable weight.^{1,5}

That comparison nevertheless needs to be stated precisely. The confidence interval permits modest harm and a potentially worthwhile benefit; it does not establish absence of any effect. The trial was designed around a comparatively large odds ratio of 1.76, not to reliably detect the observed estimate of 1.13. Its result makes the assumed large benefit unlikely in this setting but leaves smaller effects unresolved. Calling the trial simply underpowered, without specifying the effect of interest, obscures both what it excluded and what remains uncertain.^{1,5}

The correspondence focused on this issue. Goyal and colleagues argued that optimistic effect assumptions can cause useful treatments to be missed. The investigators replied that earlier thrombolysis studies supported substantial effects and emphasised the unexpectedly short opportunity for drug action before thrombectomy. Both points are relevant, but the observed odds ratio cannot be treated as an established treatment benefit or converted into a reliable number needed to treat. The confidence interval, clinical setting and consequences of additional treatment matter more than a retrospective calculation based on the point estimate.^{6,7}

Treatment timing is central to applicability. Almost all patients were enrolled where thrombectomy was immediately available, and arterial puncture followed the bolus after

a median of sixteen minutes. Clot contact occurred later, so puncture is not the same as reperfusion, but the pharmacological lead time was still short. The investigators had hoped to enrol more patients before transfer from hospitals without endovascular capability; research infrastructure and logistics prevented that. The trial therefore largely evaluated tenecteplase shortly before thrombectomy, rather than a prolonged period in which intravenous thrombolysis was the only available reperfusion treatment.^{1,5,7}

This limitation does not justify deliberately delaying thrombectomy to let tenecteplase work. Such a policy was not randomised, and any gain from drug exposure would have to be weighed against delayed definitive reperfusion. Similarly, the small actual-no-thrombectomy subgroup cannot resolve treatment for hospitals without access to the procedure. Whether thrombectomy occurred depended partly on clinical evolution, angiographic findings and technical success after allocation. Restricting analysis to that subgroup can disturb the protection provided by randomisation, while its wide interval leaves substantial uncertainty.^{1,5}

The favourable perfusion criteria also define a narrow biological population. Median cores were tiny despite treatment approximately thirteen hours after last known well. Automated imaging estimated threatened and irreversibly injured tissue; it did not prove that all mismatch tissue was viable or that the same thresholds apply without comparable acquisition and processing. TIMELESS provides little direct information about large established infarcts, patients selected only with noncontrast computed tomography, posterior-circulation occlusions or substantial prestroke dependency. The imaging requirements also impose practical constraints on generalisation to less resourced services.^{1,5}

The M1 finding is interesting but insufficient to change the primary conclusion. It arose among multiple subgroup and secondary comparisons after a neutral overall endpoint. Its interval reached the null boundary, and the study was not designed to establish a separate definitive benefit in this anatomical subgroup. The accompanying editorial discussed biological reasons why clot location might influence response. Such plausibility supports further investigation; it does not convert an exploratory estimate into a validated selection rule.^{1,5,8}

The contrast between more complete arterial recanalisation and unchanged tissue reperfusion is equally important. An open artery at twenty-four hours does not establish that flow returned early enough to preserve useful brain function. High procedural reperfusion rates may also have reduced the scope for additional benefit. Imaging was unavailable for some participants, and these comparisons were exploratory. They

support a possible pharmacological effect on vessel patency, while leaving the clinically decisive question of meaningful recovery unanswered.^{1,5}

Missing primary outcomes were relatively few, and alternative analyses were broadly consistent. However, the distinction between multiple-imputation estimates and last-observation-carried-forward displayed counts should remain explicit. An apparent mismatch between a raw proportion and an adjusted effect is not necessarily a reporting error when they arise from different methods. The three exclusions without postbaseline data mean that the primary analysis was not literally all randomised patients. Missingness was substantially greater for several ancillary neurological and quality-of-life outcomes, limiting reassurance from apparently similar secondary scores.^{1,5}

Safety also requires restraint. Seven versus five symptomatic haemorrhages cannot establish that late tenecteplase has no additional bleeding risk. The safety population excluded those not dosed and classified one crossover by treatment received, so its mortality counts differ legitimately from the efficacy display. Serious adverse events were common, and individual categories were too sparse for secure causal conclusions. A treatment without demonstrated functional benefit needs more than a nonsignificant safety comparison to justify routine use in the population studied.^{1,5}

Finally, industry involvement was substantive and transparently described. Genentech participated in design, oversight and analysis; describing the sponsor as merely providing funding would be inaccurate. Independent monitoring, masking and academic data access offered safeguards, although they are not equivalent to a wholly independent reanalysis. No single feature invalidates the findings. Taken together, the trial supports a credible neutral result for its delivered treatment strategy, with remaining uncertainty about smaller effects and materially different pathways to reperfusion.^{1,5}

Body of Evidence

Imaging-guided thrombolysis had already challenged a strictly clock-based approach. WAKE-UP randomised 503 patients with unknown onset and a diffusion-positive, FLAIR-negative magnetic resonance pattern; planned thrombectomy was excluded. A ninety-day Rankin score of zero or one occurred in 53.3% with alteplase and 41.8% with placebo, adjusted odds ratio 1.61 (1.09–2.36). This supported treatment in an imaging-defined group likely to have relatively recent infarction, rather than proving efficacy throughout a twenty-four-hour interval.²

EXTEND selected patients by perfusion mismatch for alteplase 4.5–9 hours after onset or after waking with stroke. Among 225 participants, its primary excellent-outcome endpoint occurred in 35.4% versus 29.5%, adjusted risk ratio 1.44 (1.01–2.06). Symptomatic haemorrhage occurred in 6.2% versus 0.9%, and the secondary ordinal analysis was neutral. These findings supported tissue selection while retaining uncertainty about bleeding and the size of clinical benefit. TIMELESS extended the time window and changed both the drug and the surrounding reperfusion pathway.^{1,3}

EXTEND-IA TNK concerned a different comparison: tenecteplase against alteplase before thrombectomy within 4.5 hours. In 202 participants, substantial reperfusion or absence of a retrievable thrombus at initial angiography occurred in 22% versus 10%; secondary functional outcomes also favoured tenecteplase. This provided a rationale for studying the bolus agent as bridging treatment. It could not establish its incremental value against placebo much later, especially when rapid thrombectomy followed administration.⁴

TWIST tested tenecteplase in wake-up stroke using noncontrast computed tomography selection and treatment within 4.5 hours of awakening. Among 578 patients, the adjusted odds ratio for better ninety-day function was 1.18 (0.88–1.58; $P=0.27$). Its neutral result cautioned against assuming that logistical simplicity or a late thrombolysis hypothesis alone was sufficient. Selection by an estimated tissue mismatch, selection by a magnetic resonance timing pattern and selection by noncontrast computed tomography represent distinct clinical strategies.⁹

Later in 2024, TRACE-III supplied an especially important contrast. It randomised 516 patients in China with perfusion-selected large-vessel occlusion 4.5–24 hours after last known well who lacked access to thrombectomy. Only nine underwent rescue thrombectomy. A ninety-day Rankin score of zero or one occurred in 33.0% with tenecteplase and 24.2% with standard medical treatment, relative rate 1.37 (1.04–1.81). Symptomatic intracranial haemorrhage occurred in 3.0% versus 0.8%. This study addressed a treatment setting scarcely represented in TIMELESS.¹⁰

The two 2024 trials should therefore be interpreted together without treating their results as interchangeable. Differences in thrombectomy availability, comparator care and primary outcome prevent a simple pooled clinical conclusion from these headline estimates. TIMELESS did not demonstrate better overall function when late tenecteplase was added largely before rapid thrombectomy. TRACE-III provided evidence of functional benefit, with a bleeding trade-off, where endovascular treatment was generally unavailable. Neither trial establishes that thrombectomy should be

withheld or delayed, and neither supports extending treatment to patients who lack their imaging and clinical eligibility features.^{1,10}

Does tenecteplase improve recovery 4.5–24 hours after stroke in perfusion-selected patients, most of whom undergo rapid thrombectomy?

TIMELESS did not demonstrate improved ninety-day function. Its result should not be extrapolated to patients without access to thrombectomy.

References

1. Albers GW, Juma M, Purdon B, Zaidi SF, Streib C, Shuaib A, et al.; TIMELESS Investigators. Tenecteplase for Stroke at 4.5 to 24 Hours with Perfusion-Imaging Selection. *N Engl J Med* 2024;390(8):701–711.
2. Thomalla G, Simonsen CZ, Boutitie F, Andersen G, Berthezene Y, Cheng B, et al.; WAKE-UP Investigators. MRI-Guided Thrombolysis for Stroke with Unknown Time of Onset. *N Engl J Med* 2018;379(7):611–622.
3. Ma H, Campbell BCV, Parsons MW, Churilov L, Levi CR, Hsu C, et al.; EXTEND Investigators. Thrombolysis Guided by Perfusion Imaging up to 9 Hours after Onset of Stroke. *N Engl J Med* 2019;380(19):1795–1803.
4. Campbell BCV, Mitchell PJ, Churilov L, Yassi N, Kleinig TJ, Dowling RJ, et al.; EXTEND-IA TNK Investigators. Tenecteplase versus Alteplase before Thrombectomy for Ischemic Stroke. *N Engl J Med* 2018;378(17):1573–1582.
5. TIMELESS Investigators. Supplementary appendix to: Tenecteplase for Stroke at 4.5 to 24 Hours with Perfusion-Imaging Selection. *N Engl J Med* 2024. doi:10.1056/NEJMoa2310392.
6. Goyal M, Bosshart SL, Ospel JM. Tenecteplase for Stroke at 4.5 to 24 Hours. *N Engl J Med* 2024;390(18):1729.
7. Albers GW, Purdon B, Campbell BCV. Tenecteplase for Stroke at 4.5 to 24 Hours. Reply. *N Engl J Med* 2024;390(18):1729. doi:10.1056/NEJMc2403602.
8. Leifer D. Tenecteplase for Stroke—Opening the Window? *N Engl J Med* 2024;390(8):760–761. doi:10.1056/NEJMe2314930.
9. Roaldsen MB, Eltoft A, Wilsgaard T, Christensen H, Engelter ST, Indredavik B, et al.; TWIST Investigators. Safety and efficacy of tenecteplase in patients with wake-up stroke assessed by non-contrast CT (TWIST): a multicentre, open-label, randomised controlled trial. *Lancet Neurol* 2023;22(2):117–126.
10. Xiong Y, Campbell BCV, Schwamm LH, Meng X, Jin A, Parsons MW, et al.; TRACE-III Investigators. Tenecteplase for Ischemic Stroke at 4.5 to 24 Hours without Thrombectomy. *N Engl J Med* 2024;391(3):203–212.

TRACE-III

Xiong Y, Campbell BCV, Schwamm LH, Meng X, Jin A, Parsons MW, et al.; TRACE-III Investigators. Tenecteplase for Ischemic Stroke at 4.5 to 24 Hours without Thrombectomy. N Engl J Med 2024;391(3):203–212.

Introduction

Late presentation with an occluded major cerebral artery creates a difficult decision when thrombectomy cannot be accessed. Imaging may show potentially salvageable brain tissue, but identifying that tissue is useful only if an effective reperfusion treatment can be delivered. Transfer distances, service hours and local resources can make an otherwise appropriate endovascular procedure unavailable. A rapidly administered intravenous treatment could offer an important alternative in this setting, provided that improved recovery outweighs the risk of intracranial bleeding.¹

The success of thrombectomy in DAWN and DEFUSE 3 established that selected patients can benefit from reperfusion well beyond six hours after last known well. It did not establish that an intravenous thrombolytic has the same efficacy or safety in the same time window. Mechanical clot removal and systemic fibrinolysis differ in their ability to restore flow, their speed and their potential complications. Extending the clock for one treatment therefore requires separate evidence for the other.^{2,3}

Earlier thrombolysis studies offered a rationale for this research. Imaging-selected alteplase had shown benefit beyond conventional time limits, while TRACE-2 supported tenecteplase as an alternative to alteplase within 4.5 hours. Tenecteplase's single-bolus administration is operationally attractive, but convenience does not itself justify late treatment. The key unanswered question was whether it could improve outcomes in patients with persisting anterior-circulation occlusion and favourable perfusion imaging who would otherwise receive medical treatment without reperfusion.^{1,4}

TIMELESS, published earlier in 2024, did not demonstrate better overall function with late tenecteplase, but most participants underwent rapid thrombectomy. TRACE-III deliberately addressed a different pathway: patients for whom thrombectomy was not planned and generally unavailable. Its importance lies in this distinction, rather than in a universal extension of intravenous thrombolysis to twenty-four hours after stroke.^{1,5}

Synopsis

TRACE-III was a phase 3, multicentre, open-label randomised superiority trial with blinded outcome assessment at 58 centres in China. Recruitment occurred from January 2022 to November 2023. Public research funding supported the study, and the manufacturer, China Shijiazhuang Pharmaceutical Company Recomgen Pharmaceutical, provided tenecteplase and an unrestricted infrastructure grant. The report states that the company did not participate in design, conduct or manuscript preparation and could not delay publication. The supplement lists a steering committee, data and safety monitoring board and clinical events committee.^{1,6}

Adults were eligible 4.5–24 hours after last known well, including wake-up and unwitnessed strokes. They required a prestroke modified Rankin scale score of zero or one, a National Institutes of Health Stroke Scale score of six to twenty-five, and an intracranial internal carotid or M1 or M2 middle cerebral artery occlusion on angiography. Computed tomography or magnetic resonance perfusion imaging, analysed using iStroke software, had to show a core below seventy millilitres, a hypoperfused-tissue-to-core ratio of at least 1.8, and an absolute mismatch of at least fifteen millilitres.^{1,6}

Planned thrombectomy was an exclusion. Other important exclusions included hypodensity involving more than one third of the middle cerebral artery territory on noncontrast computed tomography, intracranial haemorrhage, significant coagulopathy, recent major bleeding or procedures, and persistent systolic pressure of at least 180 mmHg or diastolic pressure of at least 100 mmHg despite treatment. Participants or their legally authorised representatives provided written consent. Central randomisation used blocks of four, with equal allocation stratified by occlusion site. Patients and treating clinicians knew assignment; certified outcome assessors and event adjudicators were masked.^{1,6}

The intervention was a single intravenous tenecteplase bolus of 0.25 mg/kg, maximum 25 mg, over five to ten seconds immediately after randomisation. Controls received antiplatelet therapy at clinicians' discretion. In the tenecteplase group, antiplatelet treatment started after twenty-four-hour imaging had excluded significant haemorrhagic transformation. Other care followed Chinese stroke guidance. Rescue thrombectomy could be performed after clinical deterioration if it became accessible; such patients remained in the intention-to-treat analysis.^{1,6}

The primary endpoint was a modified Rankin score of zero or one at ninety days: no symptoms or no significant disability despite symptoms. Secondary outcomes included the full ordinal Rankin distribution, functional independence at a score of zero to two,

major neurological improvement at seventy-two hours, and reperfusion at twenty-four hours. Recanalisation was exploratory. Safety outcomes included symptomatic intracranial haemorrhage within thirty-six hours, ninety-day mortality, systemic bleeding and adverse events. Symptomatic haemorrhage required intracranial bleeding associated with deterioration of at least four NIHSS points or death, with bleeding identified as the predominant cause.^{1,6}

The sample of 516 allowed 80% power for an increase in the primary outcome from 25% to 37%, allowing 10% attrition and a one-sided alpha of 0.025. An adaptive reassessment after 258 patients completed follow-up could have increased recruitment to 808. Its conditional power of 0.87 lay above the specified promising zone, so the target remained 516; recruitment was already close to completion when this review occurred. The primary model used log-binomial regression with a centre random effect and reported a relative rate. Secondary and subgroup intervals were not adjusted for multiplicity.^{1,6}

Of 1,469 screened patients, 516 were randomised: 264 to tenecteplase and 252 to standard medical treatment. All contributed primary outcome and safety data. Twenty-eight had major deviations and were excluded from per-protocol analyses, leaving 249 and 239. Median age was sixty-seven years overall, 67.8% were men and approximately 96% were Han Chinese. Median NIHSS score was eleven with tenecteplase and ten with control; estimated core volumes were 16.4 and 14.9 millilitres. Almost 89% were selected with computed tomography perfusion.^{1,6}

Median time from last known well to randomisation was 12.3 hours, and door-to-needle time in the tenecteplase arm was 139 minutes. Twenty-three hospitals had no thrombectomy access at all and enrolled 120 participants; participating services did not have continuous access to thrombectomy. Transfer networks and pandemic restrictions further limited access. Nine patients subsequently underwent endovascular treatment, four and five respectively: six after clinical deterioration during hours of availability and three for a recurrent stroke during the admission.^{1,6}

A ninety-day Rankin score of zero or one occurred in 87 of 264 patients given tenecteplase and 61 of 252 controls, or 33.0% versus 24.2%. The reported relative rate was 1.37 (95% confidence interval 1.04–1.81; $P=0.03$). The unadjusted absolute difference calculated from the reported counts was 8.7 percentage points. Functional independence occurred in 115/264, or 43.6%, versus 84/252, or 33.3%, relative rate 1.31 (1.05–1.63). The ordinal comparison gave a common odds ratio of 1.33 (0.98–1.81), with values above one favouring tenecteplase.¹

Major neurological improvement occurred in 40/250 patients, or 16.0%, versus 15/249, or 6.0%, relative rate 2.66 (1.51–4.69). Reperfusion exceeding 90% of the initial perfusion lesion was observed in 48/239, or 20.1%, versus 27/229, or 11.8%, relative rate 1.70 (1.10–2.64). Complete arterial recanalisation occurred in 69/247, or 27.9%, versus 14/237, or 5.9%, relative risk 4.73 (2.74–8.18). These secondary and exploratory estimates supported the primary finding but were not independently confirmatory. Analyses imputing missing secondary outcomes and the per-protocol comparison were broadly consistent.^{1,6}

Symptomatic intracranial haemorrhage occurred in eight patients given tenecteplase and two controls, or 3.0% versus 0.8%, relative rate 3.82 (0.82–17.87). Ninety-day mortality was 35/264, or 13.3%, versus 33/252, or 13.1%, relative rate 1.01 (0.65–1.58). Moderate or severe systemic bleeding occurred in five versus two patients, and serious adverse events in 53 versus 43. Central imaging review identified nine patients enrolled despite extensive territorial hypodensity; five developed symptomatic intracranial haemorrhage. The report does not provide an allocation-specific breakdown of those five events.^{1,6}

Critique

TRACE-III provides a positive randomised test of late tenecteplase in a clearly defined clinical setting. Its primary endpoint was patient-centred, follow-up was complete and the intention-to-treat analysis retained patients who subsequently underwent thrombectomy or were found to have eligibility violations. The observed control event rate was close to the planning assumption, and the trial completed its original sample. These features reduce concerns about attrition, selective postrandomisation exclusions and premature stopping after a favourable fluctuation.^{1,6}

The primary result is clinically meaningful but statistically less secure than its headline relative improvement may suggest. The absolute increase in excellent recovery was approximately nine percentage points, while the lower confidence limit for the relative rate was close to one. Replication could materially refine the estimated size of benefit. The trial was designed for a twelve-point difference and observed a smaller one; its positive result should be reported with its uncertainty, rather than described as proof of a large benefit for every eligible patient.^{1,6}

The primary binary threshold also captures only one aspect of recovery. A score of one permits symptoms without significant disability, and therefore does not mean complete neurological normality. Patients improving from severe dependence to walking with assistance may not cross this threshold despite an important benefit. Conversely, a small

change near the boundary can alter binary classification. The ordinal result favoured tenecteplase numerically but its interval included no effect. This does not negate the designated primary endpoint, although it limits claims of a conclusively demonstrated shift across the entire disability spectrum.¹

Supportive secondary findings add coherence without removing multiplicity. Functional independence, neurological improvement, reperfusion and recanalisation generally pointed in the same direction. They are correlated outcomes, however, and their unadjusted confidence intervals cannot be counted as several independent confirmations. Post hoc covariate adjustment produced more favourable functional estimates, including an ordinal interval excluding one. The published primary analysis remains the appropriate anchor; a retrospectively adjusted model should not replace it because its result looks stronger.^{1,6}

Open-label treatment is the principal design limitation. Masked assessors and a dedicated adjudication committee helped protect outcome measurement, but clinicians knew who had received a thrombolytic. Decisions about monitoring, rehabilitation, additional treatment and escalation could therefore differ. The antiplatelet timing difference was an intended part of the treatment strategies, not an accidental crossover. The evidence estimates the effect of those strategies in their clinical context, rather than an isolated drug effect under otherwise identical care.^{1,6}

Anatomical imbalance is also relevant. M2 occlusions represented 22.0% of the tenecteplase group and 15.1% of controls, despite stratification by occlusion site. Differences in baseline prognosis could influence an unadjusted comparison. The post hoc adjusted results offer some reassurance that the primary finding did not depend solely on that imbalance, but cannot prove complete control of prognostic differences. Subgroups by age, severity, timing and artery were too small to define a precise treatment-selection algorithm, and lack of a convincing interaction would not establish identical effects across them.^{1,6}

The bleeding signal deserves particular emphasis. Eight versus two symptomatic intracranial haemorrhages produced a very wide interval, compatible with a substantial increase in risk. Similar observed mortality does not establish mortality equivalence, and the trial was not powered to exclude an important survival difference. The functional endpoint incorporates death and disabling haemorrhage, so its favourable average already reflects these harms at the population level. It does not remove the need to recognise that an individual patient may be harmed by treatment.¹

The extensive-hypodensity cases make this concern practical. Five symptomatic haemorrhages occurred among nine participants who should have been excluded on that imaging criterion. This is a small, retrospectively identified subset, and the published information cannot establish which treatment caused each event or quantify a reliable subgroup treatment effect. It nevertheless shows that a favourable automated perfusion result did not ensure that all other imaging exclusions had been respected. Careful review of the noncontrast scan remains integral to reproducing the studied selection strategy.^{1,6}

Availability of thrombectomy defines the trial's relevance more strongly than geography alone. Most patients were enrolled at hospitals with some endovascular capability, but access was intermittent and transfer constrained. Only 120 came from hospitals with no access at all. It would therefore be inaccurate to describe all fifty-eight sites as unable to perform thrombectomy. Equally, patients were not randomised to tenecteplase instead of an accessible planned procedure. The study cannot establish that intravenous treatment is an adequate substitute when timely thrombectomy is feasible.^{1,6}

The accompanying editorial rightly identified the unresolved transfer pathway: late tenecteplase followed by a substantial journey for planned thrombectomy was poorly represented in both TRACE-III and TIMELESS. Their contrasting results cannot prove that longer drug exposure alone explains benefit. They differ in treatment availability, patient characteristics, outcome definitions and masking. The editorial's broader reassurance about late tenecteplase safety should also be qualified by the small haemorrhage counts and uncertainty in both trials.^{1,5,7}

Generalisation additionally depends on patient selection and infrastructure. The cohort was predominantly Han Chinese, mostly male and had moderate median stroke severity with small estimated cores. Patients with considerable prior disability, very severe deficits, posterior-circulation occlusion or large established infarction were excluded. Perfusion acquisition, automated processing, angiography and specialist scan interpretation were required. A bolus drug is simple to administer, but the full pathway is more demanding than drug availability alone suggests. The 139-minute median door-to-needle time also shows the operational burden of this trial pathway.^{1,6}

Finally, TRACE-III evaluated a specified Chinese-manufactured formulation. The supplement describes biochemical comparison with another tenecteplase product available only in abstract form at that time; this is not a direct clinical equivalence trial between formulations. Industry provided drug and infrastructure support, with reported academic control and publication freedom. These details should remain transparent without implying that funding invalidates the outcome. The strongest inference remains

improved excellent functional recovery in the selected population actually studied, accompanied by a plausible and incompletely quantified bleeding trade-off.^{1,6}

Body of Evidence

The late-reperfusion evidence begins with distinguishing treatment modalities. DAWN enrolled 206 patients 6–24 hours after last known well using a mismatch between clinical deficit and infarct volume. Thrombectomy improved both coprimary outcomes; functional independence was 49% versus 13%, with an adjusted Bayesian difference of 33 percentage points. Its early efficacy stop and selected population matter when considering magnitude, but the central result established a strong reason to pursue accessible thrombectomy in appropriately selected late presenters.²

DEFUSE 3 used perfusion selection 6–16 hours after last known well, including an estimated core below seventy millilitres and a mismatch ratio of at least 1.8. It stopped early after 182 patients and found a favourable ordinal shift with thrombectomy, odds ratio 2.77, alongside independence in 45% versus 17%. The imaging logic resembles TRACE-III, but similar eligibility thresholds do not make treatments interchangeable. Neither trial randomised patients between thrombectomy and tenecteplase.³

The earlier TRACE-2 trial addressed the choice of intravenous drug within 4.5 hours. It randomised 1,430 patients in China who were eligible for thrombolysis but not thrombectomy. Excellent recovery occurred in 439/705 evaluable tenecteplase recipients and 405/696 alteplase recipients, relative risk 1.07 (0.98–1.16), meeting its noninferiority criterion. This supported the selected tenecteplase dose in the early window. It did not demonstrate superiority over alteplase or independently establish efficacy at twelve or twenty-four hours.⁴

The TRACE-III investigators also built on earlier imaging-selected alteplase trials, particularly EXTEND, which tested a window extending to nine hours and included wake-up strokes. In their discussion, they placed TRACE-III as a further extension of tissue-based selection in patients generally not undergoing thrombectomy. That historical progression is useful, but it changes several elements simultaneously: time window, thrombolytic agent, vascular selection and clinical setting. TRACE-III did not determine whether tenecteplase was preferable to alteplase in the late window.¹

TIMELESS offers the closest 2024 comparison. It randomised 458 perfusion-selected patients 4.5–24 hours after last known well to tenecteplase or placebo, with 77.3% undergoing thrombectomy. Its primary ordinal Rankin comparison was neutral, adjusted common odds ratio 1.13 (0.82–1.57). Median bolus-to-puncture time was sixteen

minutes, and very few participants were enrolled before transfer from a hospital lacking thrombectomy capability. The trial tested an incremental treatment effect in a pathway with rapid endovascular reperfusion.⁵

TRACE-III instead showed a higher chance of excellent recovery when intravenous tenecteplase was the principal reperfusion treatment and rescue thrombectomy was rare. This is a coherent distinction between the questions asked, but it is not proof that the trials would have produced identical results if access had been exchanged. Their different primary endpoints also preclude comparing a positive binary result directly with a neutral ordinal result as though they measured precisely the same treatment effect.^{1,5}

By the end of 2024, the evidence supported considering late tenecteplase within the perfusion-selected population represented in TRACE-III when thrombectomy could not be accessed, while recognising the bleeding risk and need for careful scan review. It did not support withholding an available endovascular procedure, extending treatment to unselected late stroke, or assuming that the unresolved transfer pathway had been tested. Improving access to thrombectomy and developing an evidence-based intravenous option address related but distinct gaps in stroke care.^{1,6,7}

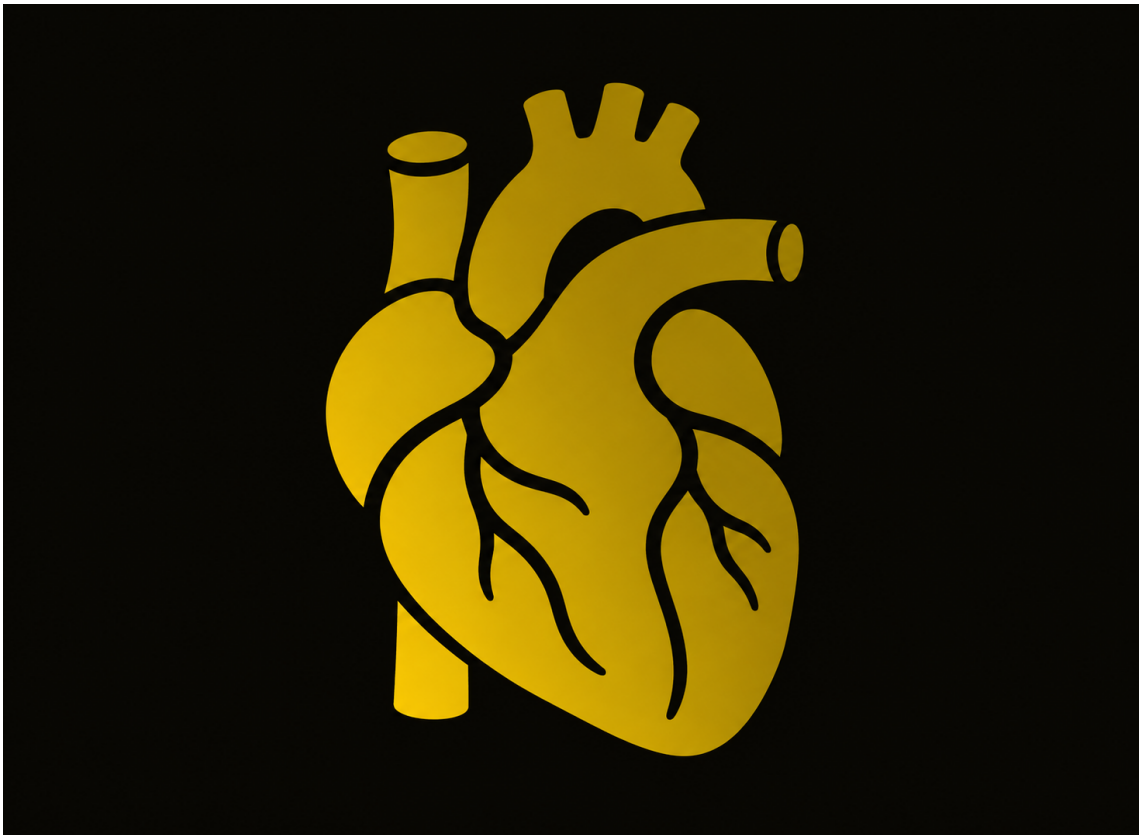
Does tenecteplase improve recovery 4.5–24 hours after stroke when perfusion imaging is favourable and thrombectomy is unavailable?

TRACE-III increased the chance of excellent ninety-day recovery, with more symptomatic intracranial bleeding. The result applies to its carefully selected population and does not justify replacing accessible thrombectomy.

References

1. Xiong Y, Campbell BCV, Schwamm LH, Meng X, Jin A, Parsons MW, et al.; TRACE-III Investigators. Tenecteplase for Ischemic Stroke at 4.5 to 24 Hours without Thrombectomy. *N Engl J Med* 2024;391(3):203–212.
2. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al.; DAWN Trial Investigators. Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct. *N Engl J Med* 2018;378(1):11–21.
3. Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, et al.; DEFUSE 3 Investigators. Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging. *N Engl J Med* 2018;378(8):708–718.

- 
4. Wang Y, Li S, Pan Y, Li H, Parsons MW, Campbell BCV, et al.; TRACE-2 Investigators. Tenecteplase versus alteplase in acute ischaemic cerebrovascular events (TRACE-2): a phase 3, multicentre, open-label, randomised controlled, non-inferiority trial. *Lancet* 2023;401(10377):645–654.
 5. Albers GW, Juma M, Purdon B, Zaidi SF, Streib C, Shuaib A, et al.; TIMELESS Investigators. Tenecteplase for Stroke at 4.5 to 24 Hours with Perfusion-Imaging Selection. *N Engl J Med* 2024;390(8):701–711.
 6. TRACE-III Investigators. Supplementary appendix to: Tenecteplase for Ischemic Stroke at 4.5 to 24 Hours without Thrombectomy. *N Engl J Med* 2024. doi:10.1056/NEJMoa2402980.
 7. Lee VH. Small Step or Giant Leap? Expanding the Acute Stroke Thrombolysis Window to 24 Hours. *N Engl J Med* 2024;391(3):273–275. doi:10.1056/NEJMe2405846.



Circulatory Trials

DanGer-Shock

Møller JE, Engstrøm T, Jensen LO, Eiskjær H, Mangner N, Polzin A, et al.; DanGer Shock Investigators. Microaxial Flow Pump or Standard Care in Infarct-Related Cardiogenic Shock. *N Engl J Med* 2024;390(15):1382–1393.

Introduction

Restoring coronary flow does not immediately reverse the circulatory failure of a large myocardial infarction. A severely impaired left ventricle may remain unable to maintain organ perfusion despite revascularisation and vasoactive treatment. Temporary mechanical support offers a plausible bridge through this period, but its physiological attraction must be balanced against invasive access, anticoagulation, blood trauma and additional intensive care. The relevant question is whether the entire treatment strategy improves outcomes that matter to patients.¹

Randomised evidence had repeatedly challenged assumptions about mechanical support. Intra-aortic balloon counterpulsation did not improve short-term survival in IABP-SHOCK II, and routine extracorporeal life support did not improve thirty-day mortality in ECLS-SHOCK. Neither finding necessarily determines the effect of a microaxial pump. These technologies provide different degrees and patterns of circulatory support, and their complications and effects on ventricular loading also differ.^{2,3}

Impella CP transfers blood from the left ventricle to the aorta, providing forward flow while unloading the ventricle. That mechanism could support recovery after coronary reperfusion, but large-bore arterial access and haemolysis could offset benefit. The small IMPRESS trial had not resolved the survival question. Observational comparisons were particularly difficult to interpret because device recipients could differ markedly in shock severity, treatment timing and the likelihood of receiving further advanced support.^{1,4}

DanGer-Shock therefore addressed an important uncertainty with a randomised comparison of routine Impella CP plus standard care against standard care alone. Its population was deliberately restricted to infarct-related shock dominated by left ventricular failure, with persistent coma after out-of-hospital cardiac arrest excluded. The trial's contribution depends on keeping this selection, its six-month survival endpoint and its substantial safety burden together when considering clinical application.¹

Synopsis

DanGer-Shock was an investigator-initiated, international, multicentre, open-label randomised trial at fourteen centres: four in Denmark, nine in Germany and one in the United Kingdom. Recruitment began in January 2013 and ended in July 2023. Slow enrolment prompted expansion from Denmark to Germany in 2019 and the United Kingdom in 2021. The Danish Heart Foundation and Abiomed funded the study. The funders had no role in design, analysis, interpretation or writing; Abiomed could review and comment on the final manuscript. An independent monitoring committee reviewed safety, and a contract research organisation monitored data accuracy and conduct.^{1,5}

Adults with ST-segment elevation myocardial infarction and cardiogenic shock were eligible. The principal report defined shock by systolic pressure below 100 mmHg or vasopressor support, arterial lactate of at least 2.5 mmol/L and left ventricular ejection fraction below 45%. Supplementary eligibility criteria also included specified electrocardiographic equivalents associated with an acute proximal coronary occlusion. Myocardial infarction had to be recent and shock could not have persisted beyond twenty-four hours. Important exclusions included overt right ventricular failure, mechanical complications of infarction, unsuitable arterial access, significant aortic valve disease and limited noncardiac life expectancy.^{1,5}

Patients remaining comatose after out-of-hospital cardiac arrest were excluded, but resuscitation in an ambulance or hospital did not itself preclude entry. Emergency consent arrangements differed between countries, using proxy or independent medical authorisation followed by patient or representative consent when possible. Web-based randomisation allocated patients equally, stratified by anterior versus nonanterior infarction and randomisation before versus after revascularisation. Allocation was unmasked. Patients could be randomised in the catheterisation laboratory or within twelve hours after leaving it, depending on when shock became apparent.^{1,5}

Impella CP was to be inserted immediately and operated at the highest possible performance for at least forty-eight hours unless complications intervened. Both groups received revascularisation and vasoactive support as indicated. Escalation to additional mechanical support was allowed in either group. Extracorporeal life support was the recommended rescue strategy in controls, although Impella 5.0 was permitted. Impella CP specifically for haemodynamic instability in a control patient was a protocol violation; its use for ventricular venting during extracorporeal support was recorded separately.^{1,5}

The primary endpoint was death from any cause at 180 days. Secondary endpoints were a composite of additional mechanical support, transplantation or death, and days alive outside hospital through 180 days. The safety composite included severe bleeding, limb

ischaemia, haemolysis, device failure or worsening aortic regurgitation. Other adverse outcomes included moderate or severe bleeding, stroke, renal replacement therapy and sepsis with positive blood cultures. Causes of death and readmission were monitored independently but were not centrally adjudicated.^{1,5}

The planned sample of 360 provided allowance for follow-up losses beyond the minimum 162 patients per group and 165 deaths needed for 80% power. Planning assumed mortality of 60% with standard care and 42% with the pump. A formal interim analysis after 180 patients completed follow-up recommended continuation; the final significance threshold was 0.048. The primary analysis used an unadjusted Cox model, with an adjusted analysis accounting for the stratification factors and country. Secondary and subgroup comparisons were not adjusted for multiplicity.^{1,5}

Of 1,211 screened patients, 360 were randomised. Five were subsequently excluded because consent could not be obtained, leaving 179 assigned to Impella CP and 176 to standard care. Primary follow-up was complete for these 355 patients. Denmark contributed 215. Median age was sixty-seven years overall, 79.2% were men, median lactate was 4.5 mmol/L and median ejection fraction was 25%. Approximately 72% had anterior infarction; 72 patients, or 20.3%, had been resuscitated before randomisation. The cohort therefore included cardiac arrest survivors despite excluding persistent coma after out-of-hospital arrest.^{1,5}

The pump was successfully placed in 170 of 179 assigned patients. Three controls received Impella CP for instability and four others received it for venting during extracorporeal support. Median time from randomisation to placement was fourteen minutes, and support continued for a median fifty-nine hours. Randomisation preceded revascularisation in 201 patients, but this did not mean that every assigned pump was inserted before the coronary intervention. Percutaneous coronary intervention was performed in 343 patients and bypass surgery in five. Among those with multivessel disease, nonculprit intervention was performed in approximately 45%.^{1,5}

At 180 days, 82/179 patients assigned to the pump and 103/176 controls had died, or 45.8% versus 58.5%. The hazard ratio was 0.74 (95% confidence interval 0.55–0.99; $P=0.04$). The observed absolute difference was 12.7 percentage points, corresponding to the investigators' point estimate of eight patients treated to prevent one death. Adjustment produced the same hazard ratio. The exploratory analysis labelled as treated gave a hazard ratio of 0.77 (0.57–1.03), after excluding patients who did not receive their assigned strategy.^{1,5}

The composite of escalation, transplantation or death occurred in 94/179 patients, or 52.5%, versus 112/176, or 63.6%, hazard ratio 0.72 (0.55–0.95). Mean days alive outside hospital were eighty-two versus seventy-three; the estimated difference was eight days (–8 to 25), providing no clear evidence of improvement. Extracorporeal support was used in 21/179 patients, or 11.7%, versus 33/176, or 18.8%. Median intensive care stay was six versus three days, and hospital stay twelve versus seven days.^{1,5}

A safety composite event occurred in 43/179 patients, or 24.0%, versus 11/176, or 6.2%. Moderate or severe bleeding occurred in 39/179, or 21.8%, versus 21/176, or 11.9%; limb ischaemia in ten versus two patients; and renal replacement therapy in 75/179, or 41.9%, versus 47/176, or 26.7%. Stroke occurred in seven versus four patients and culture-positive sepsis in twenty-one versus eight. These counts show an appreciable complication burden accompanying the survival result.¹

Critique

DanGer-Shock demonstrated a statistically significant reduction in its primary mortality endpoint after completing its planned recruitment. Randomisation, an objective endpoint and complete follow-up of the analysis cohort strengthen the causal interpretation. The study also achieved substantial treatment separation despite allowing rescue support. These are important advantages over registries, where decisions to insert a pump are closely linked to prognosis and local capability. The result deserves recognition as positive evidence for the strategy actually tested.^{1,5}

Precision nevertheless matters. The upper confidence limit of 0.99 is close to no effect, and $P=0.04$ only narrowly crossed the adjusted threshold of 0.048. The data are compatible with a substantial survival benefit but also with a much smaller effect than the observed absolute difference. The number needed to treat of eight is a point estimate, not a dependable expectation for every service or patient. Planning for an unusually large eighteen-percentage-point mortality reduction also means the trial offers less assurance about modest benefits and effect modification.¹

The five postrandomisation consent exclusions mean this was not a complete analysis of all 360 allocations, although the remaining cohort had no missing primary endpoint. Exclusions were few and their reason is transparent. The exploratory as-treated analysis does not provide a stronger causal estimate: it removes nonadherent patients and compromises randomisation. Its interval crossing one does not negate the primary analysis, just as its similar point estimate cannot independently confirm it. Both analyses should retain their proper status.^{1,5}

Open treatment allocation creates a different limitation from uncertainty about whether a patient died. Knowledge of assignment could influence escalation, ventilation, renal support and other care. The comparison consequently estimates a treatment strategy incorporating these downstream decisions. It cannot isolate the biological effect of ventricular unloading from every associated management change. More controls received extracorporeal support, while some patients in each arm progressed to durable ventricular assistance. Survival should therefore not be interpreted as evidence that all beneficiaries recovered without further advanced treatment.^{1,5}

Nor was this a randomised test of pump placement before rather than after coronary reperfusion. Randomisation timing was stratified, but insertion timing was not allocated independently. Some patients randomised before revascularisation received the device afterwards. The short median placement interval demonstrates feasibility in experienced catheterisation laboratories; it does not establish that delaying reperfusion to achieve a particular pump sequence improves outcomes. Such a recommendation would go beyond the comparison undertaken.^{1,5}

The harms are clinically substantial. About one additional patient experienced the safety composite for every six treated, using the observed absolute event difference. This cannot be subtracted from the number needed to treat for survival: the endpoints differ in severity, timing and reversibility, and the same patient may both survive because of treatment and experience a complication. Bleeding, limb ischaemia and renal support affect patients directly and require staff, equipment and procedural rescue. A survival benefit is therefore compatible with a demanding and harmful treatment course.¹

The renal analysis published later in 2024 helps interpret this trade-off. Renal replacement therapy remained more frequent with the pump after accounting for death as a competing event, subdistribution hazard ratio 1.67 (1.18–2.35). Approximately two thirds of renal replacement episodes began within twenty-four hours. All survivors were free of renal replacement therapy at 180 days, which is reassuring about persistent dialysis dependence but does not establish normal kidney health. These findings make it inappropriate to dismiss the excess renal support simply as a consequence of longer survival.⁶

Longer intensive care and hospital stays require similarly careful interpretation. Death can shorten a stay, so simple duration comparisons do not isolate resource-related harm from successful rescue. Conversely, the absence of a clear increase in days alive outside hospital limits claims that the mortality benefit was accompanied by faster overall recovery. Quality of life was not established by the principal report. The accompanying

editorial emphasised survival and the achievement of completing the trial; patient-centred implementation still requires attention to recovery and complications.^{1,7}

Selection is central to external validity. Patients with persistent coma after out-of-hospital arrest, predominant right ventricular failure, mechanical complications or unsuitable access were outside the principal target population. These exclusions may reduce competing causes of death that a left ventricular pump cannot reverse. They do not justify extending the result to all infarct-related shock, nonischaemic shock or extracorporeal resuscitation. The largely male cohort and small subgroups also prevent reliable conclusions that treatment works in one sex, age group or blood-pressure category but not another.^{1,5}

Finally, recruitment over more than ten years spanned changes in coronary and intensive care practice. Almost half of patients with multivessel disease received nonculprit intervention, including after CULPRIT-SHOCK changed the evidence for that approach. Broadly similar use between groups does not itself confound randomisation, but it affects how closely the comparator resembles a present-day local pathway. Experienced centres and substantial Danish recruitment further constrain transportability. The appropriate clinical response is selective adoption with the capacity to manage complications, rather than treating the device as a universal solution to shock.^{1,5,8}

Body of Evidence

The SHOCK trial provides the historical foundation for emergency revascularisation. It randomised 302 patients with infarction complicated by cardiogenic shock to early revascularisation or initial medical stabilisation. Thirty-day mortality, the primary endpoint, was 46.7% versus 56.0% ($P=0.11$); six-month mortality was 50.3% versus 63.1% ($P=0.027$). The distinction between the neutral primary endpoint and favourable later survival is important. SHOCK supports the relevance of sustained follow-up, but does not mean that a later positive comparison should replace an earlier prespecified primary endpoint in every trial.⁹

IABP-SHOCK II subsequently randomised 600 patients to balloon counterpulsation or control alongside planned early revascularisation. Thirty-day mortality was 39.7% versus 41.3%, relative risk 0.96 (0.79–1.17). It did not establish survival benefit from routine balloon support. DanGer-Shock does not overturn that result: a microaxial pump is a different intervention. Equally, the absence of a benefit from balloon counterpulsation was never sufficient evidence that every form of mechanical support must fail.²

IMPRESS directly compared Impella CP with an intra-aortic balloon pump, but enrolled only forty-eight mechanically ventilated patients with severe infarct-related shock. Thirty-day mortality was 46% with Impella and 50% with balloon support, hazard ratio 0.96 (0.42–2.18); six-month mortality was 50% in both groups. Such wide uncertainty could not resolve an important survival difference. The different comparator and more uniformly ventilated population also limit direct comparison with DanGer-Shock, in which only a minority were intubated at randomisation.^{1,4}

ECLS-SHOCK tested routine early extracorporeal life support in 420 randomised patients, with 417 analysed. Thirty-day mortality was 47.8% versus 49.0%, relative risk 0.98 (0.80–1.19), while moderate or severe bleeding and vascular complications were more frequent. Its one-year results, published in 2024, still showed closely similar mortality: 55.0% versus 55.8%. Thus, extending observation did not reveal an analogous late benefit for routine extracorporeal support. DanGer-Shock and ECLS-SHOCK support distinguishing devices and populations; they do not constitute a head-to-head comparison establishing Impella's superiority over extracorporeal support.^{3,10}

CULPRIT-SHOCK addressed the revascularisation strategy itself. Among 706 randomised patients with multivessel coronary disease and shock, the thirty-day composite of death or renal replacement therapy occurred in 45.9% with culprit-only intervention and 55.4% with immediate multivessel intervention, relative risk 0.83 (0.71–0.96). This matters when interpreting a pump trial recruited across an era in which nonculprit intervention remained common. DanGer-Shock cannot establish whether its absolute benefit would be identical under a consistently culprit-only initial strategy, although balanced cointervention use preserves the relevance of its internal randomised comparison.^{1,5,8}

Together, these trials discourage a single verdict on all mechanical support. Revascularisation, balloon counterpulsation, extracorporeal support and ventricular unloading are separate interventions. Their benefits depend on the disease process, treatment timing, competing causes of death and harms. DanGer-Shock adds positive survival evidence for routine Impella CP in selected STEMI-related shock, while its safety findings require substantial caution. The evidence available by the end of 2024 supports a focused clinical application, with experienced teams, appropriate access and realistic discussion of complications, rather than extrapolation to every patient with circulatory failure.^{1,7}

Should Impella CP be used routinely in patients resembling the DanGer-Shock population?

It reduced 180-day mortality in selected patients with STEMI-related cardiogenic shock, at the cost of substantially more complications. Application requires careful patient selection and an experienced team able to manage device-related harm.

References

1. Møller JE, Engstrøm T, Jensen LO, Eiskjær H, Mangner N, Polzin A, et al.; DanGer Shock Investigators. Microaxial Flow Pump or Standard Care in Infarct-Related Cardiogenic Shock. *N Engl J Med* 2024;390(15):1382–1393.
2. Thiele H, Zeymer U, Neumann FJ, Ferenc M, Olbrich HG, Hausleiter J, et al.; IABP-SHOCK II Trial Investigators. Intraaortic Balloon Support for Myocardial Infarction with Cardiogenic Shock. *N Engl J Med* 2012;367(14):1287–1296.
3. Thiele H, Zeymer U, Akin I, Behnes M, Rassaf T, Mahabadi AA, et al.; ECLS-SHOCK Investigators. Extracorporeal Life Support in Infarct-Related Cardiogenic Shock. *N Engl J Med* 2023;389(14):1286–1297.
4. Ouweneel DM, Eriksen E, Sjaauw KD, van Dongen IM, Hirsch A, Packer EJS, et al. Percutaneous Mechanical Circulatory Support Versus Intra-Aortic Balloon Pump in Cardiogenic Shock After Acute Myocardial Infarction. *J Am Coll Cardiol* 2017;69(3):278–287.
5. DanGer Shock Investigators. Supplementary appendix to: Microaxial Flow Pump or Standard Care in Infarct-Related Cardiogenic Shock. *N Engl J Med* 2024. doi:10.1056/NEJMoa2312572.
6. Zweck E, Hassager C, Beske RP, Jensen LO, Eiskjær H, Mangner N, et al.; DanGer Shock Investigators. Microaxial Flow Pump Use and Renal Outcomes in Infarct-Related Cardiogenic Shock: A Secondary Analysis of the DanGer Shock Trial. *Circulation* 2024;150(25):1990–2003.
7. Rao SV. Mechanical Circulatory Support in Cardiogenic Shock—Persistence and Progress. *N Engl J Med* 2024;390(15):1436–1438.
8. Thiele H, Akin I, Sandri M, Fuernau G, de Waha S, Meyer-Saraei R, et al.; CULPRIT-SHOCK Investigators. PCI Strategies in Patients with Acute Myocardial Infarction and Cardiogenic Shock. *N Engl J Med* 2017;377(25):2419–2432.

9. Hochman JS, Sleeper LA, Webb JG, Sanborn TA, White HD, Talley JD, et al.; SHOCK Investigators. Early Revascularization in Acute Myocardial Infarction Complicated by Cardiogenic Shock. *N Engl J Med* 1999;341(9):625–634.

10. Desch S, Zeymer U, Akin I, Behnes M, Duerschmied D, Rassaf T, et al.; ECLS-SHOCK Investigators. Routine extracorporeal life support in infarct-related cardiogenic shock: 1-year results of the ECLS-SHOCK trial. *Eur Heart J* 2024;45(39):4200–4203.

Possible Reporting Inconsistencies. The safety estimates labelled “relative risk” in Table 3 and the abstract appear to be odds ratios: 4.74 for the safety composite and 1.98 for renal replacement therapy match the crude odds ratios calculated from the published counts. The corresponding crude risk ratios are approximately 3.84 and 1.57. This appears to be an effect-measure labelling issue; the event counts and percentages above are unchanged.

Jaber WA, Gonsalves CF, Stortecky S, Horr S, Pappas O, Gandhi RT, et al.; PEERLESS Committees and Investigators. 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. Published online 29 October 2024.

Introduction

Intermediate-risk pulmonary embolism presents a difficult balance between preventing deterioration and avoiding unnecessary treatment. Patients may have substantial right ventricular dysfunction while remaining haemodynamically stable. Anticoagulation limits further thrombosis, but does not immediately remove the obstruction. Reperfusion could relieve right ventricular strain more rapidly, yet an invasive procedure or thrombolytic drug introduces its own risks. The presence of a large embolus or an abnormal scan does not by itself establish that intervention improves a patient's recovery.¹

PEITHO demonstrated this tension with systemic tenecteplase: fewer episodes of haemodynamic decompensation were accompanied by more major bleeding and stroke. Catheter-based approaches attempt to alter that balance. Local thrombolysis uses smaller drug doses delivered into the pulmonary arteries, while mechanical thrombectomy removes thrombus without requiring an infusion of thrombolytic treatment. Their procedures, access requirements, recovery pathways and potential harms differ.^{1,2}

Much of the earlier evidence for catheter intervention consisted of small studies or changes in right ventricular imaging. Improvements in these measures are useful evidence of physiological activity, but cannot substitute for reliable comparisons of clinical benefit, bleeding and longer-term function. Moreover, evidence supporting one catheter strategy against another answers a different question from whether either strategy should be added to anticoagulation.^{1,3}

PEERLESS was designed for the point at which the treating team had already decided to intervene. It compared large-bore mechanical thrombectomy using FlowTrieve with locally selected catheter-directed thrombolysis. The primary endpoint ranked serious clinical events ahead of intensive care use in a hierarchical win ratio. Understanding which components generated that result is essential: a large win ratio can reflect a major

difference in the treatment pathway without establishing a reduction in death or disabling complications.^{1,4}

Synopsis

PEERLESS was a prospective, multicentre, international, open-label randomised device trial. Recruitment occurred between February 2022 and February 2024 at fifty-seven contributing sites in the United States, Germany and Switzerland. Inari Medical, the FlowTrier manufacturer, sponsored the study and designed it jointly with the academic steering committee. The sponsor generated analysis datasets; primary and secondary analyses were performed independently by the sponsor and an external biostatistician. An independent clinical events committee adjudicated safety outcomes, and a single independent physician provided blinded central imaging review. Patients gave informed consent.^{1,4}

Eligible adults had haemodynamically stable acute pulmonary embolism, right ventricular dilatation or dysfunction, and a proximal filling defect in at least one main or lobar pulmonary artery. Symptoms had begun within fourteen days, with intervention planned within seventy-two hours of diagnosis or arrival after transfer. Initially, elevated troponin was required. An amendment broadened this to at least one additional clinical risk factor, including raised troponin, heart failure, chronic lung disease, tachycardia, lower systolic pressure, tachypnoea, hypoxaemia, syncope or raised lactate.^{1,4}

High-risk pulmonary embolism with haemodynamic instability, inability to receive therapeutic anticoagulation, right heart thrombus in transit, known chronic thromboembolic disease and life expectancy below thirty days were important exclusions. An invasively measured systolic pulmonary artery pressure of at least 70 mmHg before the therapeutic catheter entered the body was a specified postrandomisation exclusion intended to avoid chronic disease. Patients with an absolute contraindication to thrombolysis could enter a separate nonrandomised thrombectomy registry; they did not contribute to the randomised comparison.^{1,4}

Electronic randomisation allocated treatment equally, stratified by a VTE-BLEED score of at least two versus a lower score. An initial requirement that at least 35% have the higher score was removed because recruitment of such patients was slower than expected. Patients and clinicians were unmasked. The modified intention-to-treat population included randomised patients who proceeded to enrolment, defined by entry of the therapeutic catheter into the body. Follow-up assessments were scheduled at approximately twenty-four hours, discharge and thirty days.^{1,4}

The intervention was aspiration or mechanical thrombectomy with FlowTrieve. The comparator used local catheter-directed thrombolysis practice, including ultrasound-facilitated and conventional infusion catheters, with device, dose and infusion duration selected locally. Both pathways involved anticoagulation; almost all patients were already receiving parenteral anticoagulation at baseline. Location of postprocedural care was determined by the treating team. The trial contained no anticoagulation-only arm.¹

The primary endpoint was a five-level hierarchy assessed by hospital discharge or seven days after the procedure, whichever occurred first: all-cause death, intracranial haemorrhage, major bleeding, clinical deterioration or bailout treatment, and postprocedural intensive care use. The last component ranked no intensive care admission above admission for up to twenty-four hours, which ranked above a longer stay. Every thrombectomy patient was compared with every thrombolysis patient, moving down the hierarchy until a winner was identified. The win ratio divided thrombectomy wins by thrombolysis wins, excluding unresolved ties.^{1,4}

Clinical deterioration required defined sustained haemodynamic or respiratory worsening, such as prolonged hypotension, cardiac arrest, increased oxygen requirement, intubation or unplanned extracorporeal support. Bailout could also follow inadequate clinical improvement or thrombus resolution, with decisions subsequently adjudicated. A secondary win ratio omitted intensive care use. Further outcomes included individual components, twenty-four-hour symptoms and right ventricular measures, length of stay, thirty-day readmission, mortality and quality of life.^{1,4}

The minimum sample of 432 gave at least 80% power at a one-sided alpha of 0.025 for the primary endpoint, increased to 550 to allow attrition. The publication presents two-sided P values. Safety comparisons were evaluated at 0.05 without multiplicity adjustment; the supplement describes a correlation-adjusted Bonferroni approach for nonsafety secondary outcomes. A site-pooling analysis found no evidence of heterogeneity. The primary report does not provide a simple mortality effect size corresponding to its composite power calculation.^{1,4}

In total, 599 patients were randomised, 300 to thrombectomy and 299 to thrombolysis. Postrandomisation exclusions removed twenty-six and twenty-three respectively; thirty-five of these forty-nine exclusions concerned the pulmonary artery pressure criterion. The analysed groups comprised 274 and 276 patients. A further 142 entered the separate registry. Mean age was 63.7 versus 61.2 years, and 54.4% versus 51.4% were men. Raised troponin was present in 94.7% overall, 26.4% had a VTE-BLEED score of at least two, and only 4.2% had a relative thrombolytic contraindication.^{1,4}

Mean time from study-hospital presentation to catheter insertion was 22.3 versus 24.9 hours. Thrombectomy took longer in the procedure room, approximately ninety-three versus sixty-five minutes, but catheter dwell time was much shorter. In the thrombolysis group, 92.0% had bilateral catheters and 59.8% received ultrasound-facilitated treatment. Median total tissue plasminogen activator dose was 16 mg among those with available dosing data, with a median infusion duration of twelve hours per lung. The comparator therefore represented several local treatment regimens.¹

The primary win ratio favoured thrombectomy, 5.01 (95% confidence interval 3.68–6.97; $P<0.001$). Without the intensive care component, the secondary win ratio was 1.34 (0.78–2.35; $P=0.30$). Clinical deterioration or bailout occurred in 5/274 patients, or 1.8%, versus 15/276, or 5.4% ($P=0.04$). Intensive care admission occurred in 114/274, or 41.6%, versus 272/276, or 98.6%; stays exceeding twenty-four hours occurred in 53/274, or 19.3%, versus 178/276, or 64.5%.^{1,4}

There were no deaths with thrombectomy and one with thrombolysis during the primary observation window. Intracranial haemorrhage occurred in two versus one patients; nineteen patients in each arm had major bleeding, or 6.9% in both groups. At twenty-four hours, severe dyspnoea affected 13.5% versus 26.4% of assessed patients, and moderate or severe right ventricular dysfunction 42.1% versus 57.9%. Mean respiratory rate was 18.3 versus 20.1 breaths/min. Reduction in right-to-left ventricular diameter ratio was similar, 0.32 versus 0.30 ($P=0.55$). Mean total hospital stay was 4.5 versus 5.3 overnights.¹

At thirty days, the reported mortality analysis included 251 and 240 patients, with one versus two deaths. All-cause readmission occurred in 8/251, or 3.2%, versus 19/239, or 7.9% ($P=0.03$). Symptom and quality-of-life scores did not differ clearly. Device- or drug-related serious adverse events affected 34/256, or 13.3%, versus 28/244, or 11.5%. Assessment denominators varied; supplementary vital-status tracing identified additional patients alive despite missing visit assessments.^{1,4}

Critique

PEERLESS provides a substantial randomised comparison of two invasive pulmonary embolism pathways. Electronic allocation, independent event adjudication, blinded imaging review and external statistical analysis strengthen its methods. Recruitment reached the planned treated sample, and primary follow-up was complete in that analysis population. Its practical relevance is greatest when both catheter options are available and a clinical decision to intervene has already been made.^{1,4}

The primary result is statistically persuasive for the chosen composite, but its magnitude must not be interpreted as a fivefold reduction in adverse events. A win ratio summarises prioritised patient-pair comparisons, not the risk of an event in an individual patient. Although death and intracranial bleeding had highest priority, they were uncommon, leaving most pairs to be separated by lower-ranked outcomes. In the supplement, intensive care use accounted for thrombectomy wins in 59.2% of all pairs and thrombolysis wins in 6.1%, dominating the overall difference.^{1,4}

This does not make intensive care use irrelevant. Avoiding unnecessary high-acuity care may benefit patients and release scarce beds. However, it is closely tied to the intervention itself: thrombolytic infusions commonly trigger monitoring policies that differ from recovery after a completed thrombectomy procedure. Unmasked clinicians determined the care location, and the endpoint used broad categories of admission and duration. The result therefore incorporates local monitoring practice as well as any difference in clinical recovery. Its magnitude may change substantially in a service delivering catheter thrombolysis outside intensive care.^{1,4}

The secondary hierarchy excluding intensive care use did not show a clear difference, with a wide interval crossing one. That comparison is particularly useful for explaining the primary finding, although it does not prove the strategies equivalent on serious clinical outcomes. Similarly, one versus zero early deaths and three intracranial haemorrhages provide little precision for survival or neurological safety. Equal observed major bleeding rates cannot establish equivalence or exclude clinically important differences.¹

The deterioration-or-bailout result offers a potentially important signal beyond bed use, but only twenty patients experienced it and $P=0.04$ was nominal. Safety outcomes were deliberately assessed without multiplicity correction to avoid overlooking harm; that approach does not make every favourable secondary comparison independently confirmatory. Thirty-day readmission is subject to the same caution. These findings support further investigation and may inform decisions, but should not carry the certainty of the primary composite.^{1,4}

Definitions also matter. Objective deterioration occurred in four thrombectomy patients and ten thrombolysis patients; bailout occurred in one and six, with overlap in the latter group. All four thrombectomy deteriorations occurred during the procedure and resolved that day, whereas thrombolysis deteriorations included cardiac arrest and respiratory failure. Nevertheless, bailout could be triggered by persistent symptoms or inadequate thrombus resolution without meeting the stricter deterioration definition.

Independent adjudication can check justification, but cannot remove the influence of unmasked clinicians on when to obtain imaging or escalate therapy.^{1,4}

The forty-nine postrandomisation exclusions are another important qualification. Most followed a measured pressure threshold specified before catheter entry, but some reflected newly recognised contraindications, instability, withdrawal or investigator judgement. Exclusions were numerically similar between groups, which is reassuring, yet the analysis no longer preserves every randomised allocation. It estimates outcomes among patients who proceeded to therapeutic catheter insertion, rather than every patient initially assigned to a catheter strategy. No outcome should be assumed for those excluded.^{1,4}

The absence of an anticoagulation-only arm is the central limit on clinical application. Low mortality in both intervention groups cannot demonstrate benefit over medical management by comparison with historical trials. Selection, baseline prognosis, follow-up and rescue thresholds differ between studies. PEERLESS cannot identify who needs an invasive procedure, show that routine intervention improves survival, or establish that an abnormal right ventricular scan alone warrants thrombectomy. The decision to intervene remains separate from the choice between these two approaches.¹

The population also underrepresented the patients in whom avoiding thrombolytic drugs might seem most attractive. Patients with absolute contraindications entered the nonrandomised registry, while relative contraindications were uncommon in the randomised cohort. The planned minimum representation of higher bleeding risk was abandoned. Consequently, similar bleeding rates in the randomised trial should not be extrapolated to patients with recent intracranial bleeding, major surgery or other substantial contraindications. The parallel registry cannot supply a randomised comparison for them.^{1,4}

Comparator variation increases pragmatic relevance but limits device-specific conclusions. Ultrasound-facilitated treatment constituted about sixty per cent of thrombolysis procedures, with several other catheters and variable infusion regimens. PEERLESS does not establish whether ultrasound improves local thrombolysis, nor whether a shorter standardised infusion would yield the same care-use difference. Thrombectomy also required a longer procedure and fluoroscopy exposure and produced greater estimated blood loss. These trade-offs should accompany discussion of reduced intensive care use.¹

Finally, early symptomatic advantages were not accompanied by a clear difference in thirty-day quality of life. Follow-up was short and assessment denominators varied, so

prevention of chronic thromboembolic disease or durable functional improvement remains unestablished. Manufacturer involvement was substantial, although independent analysis and adjudication provide safeguards; individual data were not made available for external analysis. The most defensible conclusion is an advantage for the FlowTrieve treatment pathway in this selected comparison, driven chiefly by intensive care use, with uncertainty about harder and longer-term outcomes.^{1,4}

Body of Evidence

PEITHO remains an essential reference for the consequences of adding reperfusion to anticoagulation in intermediate-risk pulmonary embolism. Among 1,005 analysed patients, systemic tenecteplase reduced seven-day death or haemodynamic decompensation from 5.6% to 2.6%, odds ratio 0.44 (0.23–0.87). Major extracranial bleeding increased from 1.2% to 6.3%, and stroke from 0.2% to 2.4%. Thirty-day mortality did not differ significantly. This established that preventing deterioration could come at a serious cost; it did not establish that catheter intervention produces the same balance.²

ULTIMA tested ultrasound-assisted catheter thrombolysis plus heparin against heparin alone in fifty-nine patients. Its primary endpoint was a change in the right-to-left ventricular diameter ratio at twenty-four hours, with a mean reduction of 0.30 versus 0.03. No major bleeding occurred, but the study was too small to provide strong reassurance about uncommon harms or survival. It supplied randomised evidence of faster reversal of right ventricular dilatation, while leaving the central clinical benefit question unresolved.³

CANARY compared conventional catheter thrombolysis with anticoagulation alone and was stopped early during the COVID-19 pandemic after ninety-four patients. Its primary three-month imaging endpoint occurred in 2/46 versus 5/39 patients, odds ratio 0.31 (0.06–1.69; $P=0.24$). Some secondary imaging measures favoured intervention, and one nonfatal major bleed occurred in that group. The incomplete, small trial supported a larger clinical-outcome study rather than a definitive recommendation for routine catheter treatment.⁵

SUNSET sPE addressed a different question within catheter thrombolysis: whether adding ultrasound improved pulmonary arterial thrombus reduction compared with a conventional infusion catheter. It did not demonstrate that advantage. This is relevant to PEERLESS because mixing ultrasound-facilitated and conventional catheter treatment cannot simply be assumed to have weakened its comparator. Equally, failure to

demonstrate superiority for one imaging endpoint is not proof that every catheter, dose or infusion duration has identical clinical effects.⁶

These studies illustrate why the sequence of clinical questions matters. Trials comparing catheter treatment with anticoagulation ask whether intervention adds benefit. Trials comparing two catheter strategies ask which pathway performs better after intervention has been chosen. Imaging endpoints test physiological effects; a hierarchy dominated by intensive care use tests a broader but still partly operational outcome. Evidence at one of these levels cannot automatically answer the others. PEERLESS advances the comparative-intervention question while leaving the initial treatment threshold unresolved.^{1,3,5}

Its early symptom and right ventricular findings are compatible with more rapid relief after mechanical clot removal, but do not prove that this mechanism caused fewer deteriorations or readmissions. Some physiological comparisons were similar, and the thirty-day patient-reported outcomes did not show sustained separation. The main report identified ongoing randomised comparisons with anticoagulation as the next step. Their results were not available within this book's 2024 evidence horizon and should not be anticipated.¹

For practice, PEERLESS can inform a multidisciplinary discussion when an invasive procedure is already considered appropriate, with local expertise, bleeding risk and monitoring arrangements made explicit. It offers a reason to consider large-bore thrombectomy over the catheter-thrombolysis pathways tested, especially where prolonged infusion mandates intensive care. It does not provide evidence for routine thrombectomy in all intermediate-risk pulmonary embolism, and it does not demonstrate a survival advantage or a reduction in major bleeding.^{1,4}

Does PEERLESS support choosing large-bore thrombectomy over catheter-directed thrombolysis?

It favoured thrombectomy mainly through reduced intensive care use, with fewer deterioration or bailout events. It did not establish better survival, less major bleeding or superiority over anticoagulation alone.

References

1. Jaber WA, Gonsalves CF, Stortecky S, Horr S, Pappas O, Gandhi RT, et al.; PEERLESS Committees and Investigators. 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. Published online 29 October 2024.
2. Meyer G, Vicaut E, Danays T, Agnelli G, Becattini C, Beyer-Westendorf J, et al.; PEITHO Investigators. Fibrinolysis for Patients with Intermediate-Risk Pulmonary Embolism. *N Engl J Med* 2014;370(15):1402–1411.
3. 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.
4. PEERLESS Committees and Investigators. Supplemental material to: Large-Bore Mechanical Thrombectomy Versus Catheter-Directed Thrombolysis in the Management of Intermediate-Risk Pulmonary Embolism. *Circulation*. Published online 29 October 2024. doi:10.1161/CIRCULATIONAHA.124.072364.
5. Sadeghipour P, Jenab Y, Moosavi J, Hosseini K, Mohebbi B, Hosseinsabet A, et al. Catheter-Directed Thrombolysis vs Anticoagulation in Patients With Acute Intermediate-High-Risk Pulmonary Embolism: The CANARY Randomized Clinical Trial. *JAMA Cardiol* 2022;7(12):1189–1197.
6. Avgerinos ED, Jaber W, Lacomis J, Markel K, McDaniel M, Rivera-Lebron BN, et al.; SUNSET sPE Collaborators. Randomized Trial Comparing Standard Versus Ultrasound-Assisted Thrombolysis for Submassive Pulmonary Embolism: The SUNSET sPE Trial. *JACC Cardiovasc Interv* 2021;14(12):1364–1373.

BLENDER

Burrell A, Bailey MJ, Bellomo R, Buscher H, Eastwood G, Forrest P, et al.; BLENDER Trial Investigators. Conservative or liberal oxygen targets in patients on venoarterial extracorporeal membrane oxygenation. *Intensive Care Med* 2024;50(9):1470–1483.

Introduction

Venoarterial extracorporeal membrane oxygenation restores circulation and oxygen delivery in patients with profound cardiac failure or refractory cardiac arrest. The same circuit can expose the arterial circulation to very high oxygen tensions. Oxygen delivery depends on blood flow and haemoglobin as well as saturation, so raising oxygen tension far above the point of haemoglobin saturation may add relatively little useful oxygen content. It may nevertheless increase oxidative stress, vasoconstriction and reperfusion injury. Observational associations between hyperoxaemia and poor outcomes made oxygen titration an attractive intervention to test.¹

The opposing risk is insufficient oxygenation. Patients receiving venoarterial support may have severely impaired lung function and an unstable native circulation. Blood leaving the oxygenator mixes with blood ejected by the heart, and the oxygen supply to different arterial territories can differ. Reducing oxygen to the circuit or ventilator therefore requires monitoring of both the patient and the extracorporeal blood. A satisfactory measurement from one location cannot describe every organ's oxygen exposure.^{1,2}

Randomised trials outside extracorporeal support had already shown that physiological plausibility was an unreliable guide to clinical benefit. Reducing supplemental oxygen did not improve ventilator-free days in ICU-ROX, while other trials raised concern about more restrictive targets in particular populations. These comparisons could not resolve the specific problem of venoarterial support, where the oxygenator provides a second, independently adjustable source of oxygen.^{1,3}

BLENDER directly compared conservative and liberal oxygen strategies in adults receiving venoarterial extracorporeal membrane oxygenation. Its registry-based design made randomisation possible within a relatively uncommon, resource-intensive treatment. The clinical question was whether reducing oxygen exposure improved time alive outside intensive care, without introducing hypoxic complications. Interpretation requires attention to the oxygen separation achieved, the frequent departures from the

conservative strategy and the suitability of an outcome whose median was zero in both groups.^{1,2}

Synopsis

BLENDER was a multicentre, open-label, parallel-group randomised trial embedded in the EXCEL extracorporeal support registry. Thirteen Australian hospitals screened patients and twelve recruited between September 2019 and June 2023. Funding came from Australia's Medical Research Future Fund. Ethics approval permitted initial enrolment without prospective consent because of the emergency setting; patients or surrogates were subsequently approached, and all agreed to continue. An independent monitoring committee reviewed safety after 150 patients had completed sixty-day follow-up.^{1,2}

Adults older than eighteen years commencing venoarterial support were eligible if randomisation could occur within six hours. Exclusions included known pregnancy, unwillingness to receive blood products, participation in another oxygen-titration study and a clinician requirement for a particular oxygen target. The trial included both cardiogenic shock and extracorporeal cardiopulmonary resuscitation for refractory cardiac arrest. It did not study patients receiving venovenous support for isolated respiratory failure.¹

A secure web system assigned patients equally to conservative or liberal oxygenation using variable blocks, stratified by hospital and indication for support. Bedside clinicians knew allocation; outcome assessors, analysing statisticians and authors were masked to group identity. The intervention continued until extracorporeal support stopped or sixty days after randomisation, whichever occurred first. Other management followed local practice, with no restrictions on concomitant care.^{1,2}

In the conservative group, oxygen supplied to the oxygenator was initially reduced to a fraction of 0.60, then adjusted between 0.50 and 1.00 to target post-oxygenator saturation of 92–96%. Ventilator oxygen was subsequently titrated, with a minimum fraction of 0.21, to a right-arm saturation of 92–96%. In the liberal group, oxygenator oxygen remained at 1.00, while ventilator oxygen targeted right-arm saturation of 97–100% and could not be reduced below 0.50. Nonventilated patients received titrated nasal or facemask oxygen. Both groups had standardised responses to hypoxia, including immediate increases in oxygen during emergencies.^{1,2}

Monitoring combined right-hand pulse oximetry, patient arterial blood gases and post-oxygenator blood sampling at least twice daily. The registry captured blood-gas

values approximately twelve-hourly during the first seven days, although clinical sampling was more frequent when required. Oxygen management and compliance monitoring continued during longer runs, but later blood-gas values were not centrally recorded. Registry data were linked to trial-specific records and the Australian and New Zealand Intensive Care Society Adult Patient Database.^{1,2}

The primary outcome was days alive and free of intensive care between randomisation and day twenty-eight. Patients dying before day twenty-eight received zero days. Secondary outcomes included ICU-free days to day sixty, mortality through day 180, duration of extracorporeal support and ventilation, lengths of stay, and six-month disability and quality of life. A sensitivity analysis replaced actual discharge with the time patients were considered ready to leave intensive care. Safety surveillance included critical hypoxia requiring an emergency response, resuscitation and seizures.¹

Pilot data suggested a mean of 11.8 ICU-free days with a standard deviation of 8.4. Detecting a three-day difference with 80% power and a two-sided alpha of 0.05 required 124 patients per group; this was increased to 300 overall for nonnormality and possible attrition. The primary intention-to-treat analysis compared medians using quantile regression. Adjustment included hospital, indication, age, diagnosis and time from cannulation to randomisation. Duration outcomes used competing-risk methods accounting for death. Secondary and subgroup comparisons were not adjusted for multiplicity.^{1,2}

During recruitment, 934 patients receiving venoarterial support were recorded in EXCEL, including 594 at participating hospitals. Three hundred entered BLENDER: 149 conservative and 151 liberal. All contributed to the primary analysis. Cardiogenic shock accounted for 192 patients and refractory arrest for 108. Mean age was approximately fifty-three years, and two-thirds were men. Acute myocardial infarction, perioperative support, decompensated heart failure and pulmonary embolism were prominent diagnoses. Mean time from support initiation to randomisation was 3.0 versus 2.5 hours.^{1,2}

The strategies separated oxygen exposure. Patient and post-oxygenator tensions and saturations were lower with conservative treatment. At least one patient arterial oxygen tension above 300 mmHg occurred in 40/149 patients, or 26.8%, versus 65/151, or 43.1%. However, only 46.5% of recorded patient saturation measurements were within the conservative target, compared with 77.2% within the liberal target. Post-oxygenator tensions below 60 mmHg occurred in 43/1,013 samples, or 4.2%, versus 3/909, or 0.3%. Blood flow, haemoglobin, lactate and use of renal support were similar.^{1,2}

Major deviations affected forty-four conservative patients, or 29.5%, and two liberal patients, or 1.3% ($P<0.001$). This category included repeated periods outside the allocated oxygen-management strategy, clinician withdrawal or conversion to another support mode. Eleven versus two patients were withdrawn from the intervention by clinicians; these categories overlapped. All remained in the intention-to-treat analysis. The per-protocol population consequently contained 105 conservative and 149 liberal patients.^{1,2}

Median ICU-free days to day twenty-eight were zero in both groups, with interquartile ranges of 0–13.7 and 0–13.3. The unadjusted median difference was zero days (95% confidence interval –3.1 to 3.1; $P=1.00$); the adjusted difference was also zero (–2.0 to 2.0). Accounting for discharge delay did not materially change the result. Median ICU-free days to day sixty were 19.3 versus 22.7, an unadjusted difference of –3.4 days (–19.0 to 12.1). Exploratory subgroup and per-protocol analyses showed no clear treatment effect.¹

By day twenty-eight, 59/149 conservative patients, or 39.6%, and 59/151 liberal patients, or 39.1%, had died, relative risk 1.01 (0.77–1.34). Day-sixty mortality was 64/149, or 43.0%, versus 62/151, or 41.1%. At day 180, with vital status available for 290 patients, mortality was 68/143, or 47.6%, versus 69/147, or 46.9%. Median support duration was 4.1 versus 4.2 days and intensive care stay 11.8 versus 11.2 days; competing-risk analyses showed no clear separation.¹

Six-month functional assessments were available for seventy-three and sixty-seven patients. Median raw WHODAS disability scores were seven in each group, and median EQ-5D-5L utility scores were 0.93 versus 0.91, without a clear difference. Critical hypoxia events occurred in sixteen patients, or 10.7%, versus thirteen, or 8.6%; resuscitation occurred in three versus six and clinical seizures in two per group. These small event counts did not demonstrate a safety difference.¹

Critique

BLENDER addresses a practical intervention directly rather than extrapolating from oxygen trials in patients without extracorporeal support. Concealed electronic allocation, stratification by indication and hospital, complete primary follow-up and an intention-to-treat analysis support internal validity. Recruitment reached its planned size. Embedding the trial in established registries also allowed comparison with patients treated outside the study, an advantage over trials that provide little information about the population from which participants were selected.^{1,2}

The central limitation is the primary endpoint's distribution. Death and remaining in intensive care both produce zero ICU-free days. With substantial early mortality and prolonged stays, at least half of each group accumulated no free days, fixing both medians at the floor. Differences among patients elsewhere in the distribution need not move that median. A P value of 1.00 therefore does not indicate identical outcomes for every patient or exceptional precision about all clinically important effects.¹

The power calculation used a three-day difference and pilot data with considerably more free days than occurred in BLENDER. Observed mean values were 5.9 and 6.4 days, despite a standard deviation similar to that anticipated. The larger representation of refractory cardiac arrest contributed to the less favourable outcome distribution. Reaching 300 participants cannot by itself ensure the intended sensitivity when the endpoint behaves differently from the planning assumptions. The unadjusted interval still includes a difference of approximately three days in either direction.^{1,2}

Mortality and longer-term outcomes provide useful additional information, but do not establish equivalence. The day-twenty-eight mortality interval is compatible with appreciable benefit or harm, and the study was not designed around a mortality difference or noninferiority margin. The reliable interpretation is that no convincing clinical advantage emerged, with uncertainty remaining about smaller effects and uncommon harms.¹

The achieved exposure contrast was real, especially in post-oxygenator oxygen tension. It would be wrong to dismiss the study as a comparison of indistinguishable oxygenation. Nevertheless, conservative treatment did not eliminate hyperoxaemia, and fewer than half of patient saturation measurements were within its target. The minimum oxygenator fraction of 0.50 sometimes prevented further reduction despite an above-target saturation. BLENDER consequently evaluates these implementable treatment algorithms, not complete avoidance of hyperoxaemia or an unrestricted search for the lowest possible oxygen delivery.^{1,2}

Adherence was markedly asymmetric. A major deviation in almost thirty per cent of the conservative group indicates that the lower-target pathway was substantially harder to sustain. This may reflect physiological instability, clinician concern, operational difficulty or a combination; the available data do not establish the reasons in individual patients. Departure from treatment could reduce both its intended effect and its risks. The intention-to-treat estimate appropriately includes that response to the strategy as deployed in practice.^{1,2}

The per-protocol result cannot settle whether perfect adherence would work. Removing forty-four patients from one arm and two from the other creates groups selected partly by events and decisions occurring after randomisation. These patients may differ in prognosis, and the balanced allocation no longer protects the comparison fully. Moreover, the same median endpoint remains vulnerable to a floor effect. Consistency with the main analysis is reassuring but is not an unbiased estimate of the biological effect of sustained lower oxygenation.¹

Monitoring limitations also affect mechanistic interpretation. Twice-daily registry entries cannot quantify every transient episode, its duration or the cumulative exposure between samples. Approximately one-quarter of patients had support extending beyond the centrally recorded oxygenation period. The side of radial arterial sampling was not consistently known, although right-hand pulse oximetry was specified. In a circulation receiving blood from both the native heart and an extracorporeal circuit, this limits conclusions about regional oxygenation and the exposure of particularly vulnerable organs.^{1,2}

The safety findings require similarly measured language. Critical hypoxia requiring emergency intervention was not clearly more frequent with conservative treatment, but post-oxygenator samples below 60 mmHg were more common. A narrowly defined emergency event is not a complete measure of potentially harmful hypoxaemia. Small numbers of seizures or resuscitations cannot exclude meaningful differences, particularly when clinicians could increase oxygen or withdraw a patient from the assigned strategy. Rescue responses are part of the tested approach, not evidence that oxygen restriction is intrinsically harmless.^{1,2}

Open treatment assignment could influence decisions about concomitant care, withdrawal and discharge. Masked outcome assessment and the discharge-readiness sensitivity analysis reduce some concerns, while similar measured flow, haemoglobin and lactate provide reassurance about major accompanying physiological differences. They do not exclude unmeasured cointerventions. Registry efficiency came with limited detail about vasopressor doses, biochemical injury and reasons for deviations; these omissions restrict explanation of a neutral result rather than invalidate randomisation.¹

Generalisability is strongest to experienced Australian services able to implement both algorithms promptly. Only 300 of 594 patients treated at participating hospitals were recruited. Demographic similarity to other registry patients is helpful, but participants had lower mortality, and missed enrolment or a clinician-mandated target could select a more manageable population. The six-hour eligibility window also excludes many late

transfers. Findings cannot automatically be extended to children, venovenous support or centres with different staffing and monitoring resources.^{1,2}

Finally, the trial combines heterogeneous causes of shock with refractory arrest. Stratification preserved balance for that distinction, but subgroup sizes were insufficient to establish separate optimal targets. Follow-up beyond discharge is a strength, although six-month functional results concern available survivors and cannot describe recovery in everyone randomised. BLENDER weakens the claim that routinely lowering oxygen in this way produces a substantial clinical benefit. It does not prove that hyperoxaemia is biologically innocent or that more extreme oxygen exposures are acceptable.¹

Body of Evidence

ICU-ROX randomised 1,000 mechanically ventilated adults to conservative or usual oxygen. Both strategies used a default lower saturation limit of 90%; conservative care also limited higher saturation and reduced oxygen when possible. Ventilator-free days did not improve, with an estimated difference of –0.3 days (95% confidence interval –2.1 to 1.6). This showed that reducing exposure need not improve recovery, but its ventilator-only setting could not answer how an extracorporeal oxygenator should be managed.³

LOCO2 tested lower targets in acute respiratory distress syndrome: arterial oxygen tension of 55–70 versus 90–105 mmHg. It stopped after 205 enrolments for safety concerns and futility. Among 201 analysed patients, twenty-eight-day mortality was 34.3% versus 26.5%, difference 7.8 percentage points (–4.8 to 20.6); the lower-target arm also had five mesenteric ischaemic events and a concerning ninety-day mortality signal. Early stopping limits precision, but the study cautions against assuming that progressively lower targets must be beneficial.⁴

BOX provided a larger comparison after out-of-hospital cardiac arrest, using arterial oxygen targets of 68–75 versus 98–105 mmHg. Among 789 analysed patients, death or hospital discharge with severe disability or coma within ninety days occurred in 32.0% versus 33.9%, hazard ratio 0.95 (0.75–1.21). This was compatible with no important advantage from the lower range, while leaving uncertainty around smaller effects. Patients had regained circulation; the comparison was not a trial of oxygen management during extracorporeal resuscitation.⁵

EXACT intervened earlier, beginning oxygen titration after prehospital return of spontaneous circulation and continuing until intensive care admission. It compared

saturations of 90–94% with 98–100% and stopped early during the pandemic. Survival to discharge was 82/214, or 38.3%, versus 101/211, or 47.9%, an absolute difference of –9.6 percentage points (–18.9 to –0.2). The study did not support the lower target. Its early setting and monitoring challenges distinguish it from controlled oxygen titration during established extracorporeal support.⁶

These trials should not be grouped simply into “conservative” and “liberal” camps. They altered different oxygen sources, used different saturation or tension ranges, began at different stages of illness and selected patients with different risks of ischaemia. A target described as liberal in one study may be far below the post-oxygenator tension produced by BLENDER's fixed oxygenator fraction of 1.00. Conversely, its conservative patient saturation range was less restrictive than that used in some non-ECMO studies. The actual exposure and clinical setting matter more than the label.^{1,2}

The observational rationale for BLENDER remains relevant but cannot override its randomised findings. Hyperoxaemia may mark low native cardiac output and severe underlying disease, as well as potentially contribute to injury. Randomisation tests whether changing an oxygen-management strategy improves outcomes; it does not isolate every biological mechanism linking oxygen tension to prognosis. The combination of exposure separation and neutral clinical results makes a large benefit from this particular algorithm less persuasive, without proving that all observational associations are noncausal.¹

For practice, the trial supports caution about mandating the tested conservative strategy solely to improve survival or shorten intensive care treatment. Lower oxygen exposure was achievable, but maintaining the pathway was more difficult and no patient-centred benefit was demonstrated. Adequate monitoring and prompt correction of hypoxia remain integral to either strategy. The unresolved question is whether a different, reliably delivered oxygen approach can improve meaningful recovery in a sufficiently large, well-characterised venoarterial support population.^{1,2}

Does conservative oxygenation improve outcomes during venoarterial ECMO?

BLENDER found no improvement in ICU-free days or other clinical outcomes with the tested conservative strategy. It caused more treatment deviations, and the trial does not establish equivalence or exclude smaller benefits and harms.

References

1. Burrell A, Bailey MJ, Bellomo R, Buscher H, Eastwood G, Forrest P, et al.; BLENDER Trial Investigators. Conservative or liberal oxygen targets in patients on venoarterial extracorporeal membrane oxygenation. *Intensive Care Med* 2024;50(9):1470–1483.
2. BLENDER Study Investigators. Supplementary appendix to: Conservative or liberal oxygen targets in patients on venoarterial extracorporeal membrane oxygenation. *Intensive Care Med* 2024. doi:10.1007/s00134-024-07564-8.
3. Mackle D, Bellomo R, Bailey M, Beasley R, Deane A, Eastwood G, et al.; ICU-ROX Investigators and the Australian and New Zealand Intensive Care Society Clinical Trials Group. Conservative Oxygen Therapy during Mechanical Ventilation in the ICU. *N Engl J Med* 2020;382(11):989–998.
4. Barrot L, Asfar P, Mauny F, Winiszewski H, Montini F, Badie J, et al.; LOCO2 Investigators and REVA Research Network. Liberal or Conservative Oxygen Therapy for Acute Respiratory Distress Syndrome. *N Engl J Med* 2020;382(11):999–1008.
5. Schmidt H, Kjaergaard J, Hassager C, Mølstrøm S, Grand J, Borregaard B, et al. Oxygen Targets in Comatose Survivors of Cardiac Arrest. *N Engl J Med* 2022;387(16):1467–1476.
6. Bernard SA, Bray JE, Smith K, Stephenson M, Finn J, Grantham H, et al.; EXACT Investigators. Effect of Lower vs Higher Oxygen Saturation Targets on Survival to Hospital Discharge Among Patients Resuscitated After Out-of-Hospital Cardiac Arrest: The EXACT Randomized Clinical Trial. *JAMA* 2022;328(18):1818–1826.



Resuscitation Trials

PARAMEDIC-3

Couper K, Ji C, Deakin CD, Fothergill RT, Nolan JP, Long JB, et al.; PARAMEDIC-3 Collaborators. A Randomized Trial of Drug Route in Out-of-Hospital Cardiac Arrest. N Engl J Med 2025;392(4):336–348. Published online 31 October 2024.

Introduction

Vascular access during cardiac arrest must be obtained while chest compressions, ventilation and defibrillation continue. Peripheral veins can be difficult to cannulate in a poorly perfused patient, particularly in confined surroundings. Intraosseous access offers a practical alternative: insertion into the marrow cavity can provide a route for resuscitation drugs when a vein is inaccessible. The unresolved question was whether its technical convenience justified making it the first approach for everyone requiring drug treatment.¹

The proposed benefit depended on more than inserting a needle successfully. Earlier access would need to produce earlier effective drug delivery, and drugs administered into bone would need to reach the circulation adequately during cardiopulmonary resuscitation. Insertion site, perfusion and displacement could all matter. Conversely, repeated intravenous attempts might delay treatment. Observational comparisons were difficult to interpret because intraosseous access could be selected after failed intravenous attempts or in patients with especially prolonged arrests.^{1,2}

Drug efficacy also provides an important distinction between process and outcome. PARAMEDIC2 established a thirty-day survival advantage from adrenaline, but did not demonstrate a clear improvement in favourable neurological survival. Obtaining access or restoring circulation is therefore not enough to establish meaningful recovery. A route trial needed patient-centred follow-up rather than technical success alone.³

PARAMEDIC-3 tested an intraosseous-first against an intravenous-first strategy across UK emergency medical services. It asked whether starting with intraosseous access improved thirty-day survival in adults with ongoing out-of-hospital cardiac arrest who needed drugs. Its size and pragmatic setting offered an opportunity to move beyond confounded route comparisons. Interpretation nevertheless depends on the timing actually achieved, rescue access after unsuccessful attempts, and the consequences of recruitment ending well before the intended sample size.^{1,2}

Synopsis

PARAMEDIC-3 was a multicentre, open-label, superiority trial conducted in ten NHS ambulance services and one air ambulance service in England and Wales. Recruitment ran from 11 November 2021 to 1 July 2024. The University of Warwick sponsored the study, with funding from the National Institute for Health and Care Research. The funder had no role in trial design, analysis or reporting. Independent steering and data monitoring committees provided oversight, and patients and members of the public contributed to the study.^{1,2}

Adults aged eighteen years or older with out-of-hospital cardiac arrest were eligible when attended by a trial-trained paramedic and vascular access was required for drug administration during continuing resuscitation. Pregnancy and existing vascular access were exclusions. Patients requiring a particular route could not undergo a meaningful comparison. The emergency setting precluded prospective patient consent; an approved deferred process allowed survivors or their representatives to decide about continued participation.^{1,2}

Sequentially numbered, sealed, tamper-evident packs contained the allocation, with equal assignment stratified by emergency medical service. Paramedics opened a pack when the patient was eligible. Neither treating clinicians nor outcome assessors were masked. Training covered trial procedures; additional vascular-access skills training was not required because both techniques were already part of routine paramedic practice.^{1,2}

The assigned strategy governed the first two vascular-access attempts. After two unsuccessful attempts, the paramedic could use either route according to clinical judgement. Choice of insertion site was also discretionary. Participating services used the EZ-IO system for intraosseous access, although the study did not require a particular manufacturer's device. Resuscitation drugs were to use the established route until circulation returned, resuscitation ended, access was lost or the patient reached hospital. Subsequent care followed usual practice.^{1,2}

The primary outcome was survival thirty days after randomisation. Secondary outcomes included any return of spontaneous circulation, sustained circulation to hospital handover, survival to discharge and favourable neurological outcome at discharge, defined as a modified Rankin Scale score of three or less. This threshold includes moderate disability. Longer-term outcomes and economic analyses were planned but were not part of this initial report. Survival information was obtained from ambulance and hospital records and national tracing sources.^{1,2}

The intended sample of 15,000 provided 90% power to detect an increase in thirty-day survival from 3.2% to 4.2%, using a two-sided significance level of 0.05. Interim efficacy analyses were planned after approximately 10% and 50% of recruitment, with O'Brien–Fleming boundaries. Funding ended before recruitment was complete. The decision to stop followed discussion with the sponsor and steering committee while investigators remained blinded to comparative outcomes; it was not a declaration of futility or evidence that the strategies were equivalent.^{1,2}

The primary analysis followed assigned strategy, excluding erroneous randomisations. Logistic regression adjusted for age, sex, witness status, bystander cardiopulmonary resuscitation, initial rhythm, arrest cause and time from emergency call to drug administration. Unadjusted, missing-data, weighting and per-protocol analyses examined robustness. Additional post hoc models removed drug timing or replaced it with response time. Secondary comparisons were not protected by adjustment for multiplicity.^{1,2}

Of 10,723 patients screened, 6,096 were randomised equally. Existing vascular access accounted for 3,914 exclusions from screening, or 36.5% of those screened. Fourteen erroneous randomisations were removed, leaving 3,040 intraosseous-first and 3,042 intravenous-first patients. Thirty-day survival was available for 3,030 and 3,034, respectively: 99.7% of the analysis population. Recruitment reached approximately 41% of its intended size, before the second planned interim analysis.^{1,2}

Patients were typically in their late sixties; approximately 64% were men and 79% arrested at home. Around 70% received bystander cardiopulmonary resuscitation. Initial shockable rhythms occurred in 18.6% versus 20.8%. Most arrests had a medical cause, although small numbers followed trauma. Median emergency response time was eight minutes in both groups.^{1,2}

First-attempt access succeeded in 2,654/3,040 patients, or 87.3%, versus 1,646/3,042, or 54.1%. The first successful route was intraosseous in 94.4% of the intraosseous-first arm and 32.6% of the intravenous-first arm. Most intraosseous-first patients received proximal tibial access; proximal humeral insertion was less common. Defined crossovers—use of the other route before two unsuccessful allocated attempts—occurred in 127 patients, or 4.2%, versus 401, or 13.2%. Permitted rescue access after two failures did not constitute crossover.^{1,2}

Median time from the emergency call to successful access was twenty-one versus twenty-two minutes. More importantly, median call-to-drug time was twenty-four minutes in both groups, with interquartile ranges of nineteen to thirty and twenty to

thirty-one minutes. Adrenaline was administered to 94.3% versus 93.2%, with a median total dose of 5 mg in each arm. Thus, markedly better first-attempt success did not produce a corresponding reduction in the median time to drug administration.^{1,2}

At thirty days, 137/3,030 intraosseous-first patients, or 4.5%, and 155/3,034 intravenous-first patients, or 5.1%, were alive. The adjusted odds ratio was 0.94 (95% confidence interval 0.68–1.32; $P=0.74$). A post hoc unadjusted risk difference was –0.6 percentage points (–1.7 to 0.5); the adjusted difference was –0.2 points (–1.1 to 0.8). Sensitivity analyses did not demonstrate a survival advantage.^{1,2}

Any return of spontaneous circulation occurred in 1,092/3,031 patients, or 36.0%, versus 1,186/3,035, or 39.1%, adjusted odds ratio 0.86 (0.76–0.97). Sustained circulation to hospital handover occurred in 654/3,016, or 21.7%, versus 744/3,023, or 24.6%, adjusted odds ratio 0.85 (0.74–0.98). Both estimates favoured the intravenous-first strategy, although these were secondary outcomes.^{1,2}

Survival to discharge was 112/3,012, or 3.7%, versus 120/3,012, or 4.0%, without a clear adjusted difference. Favourable neurological outcome occurred in 80/2,994 patients, or 2.7%, versus 85/2,986, or 2.8%, adjusted odds ratio 0.91 (0.57–1.47). Fifty-eight patients remained in hospital when discharge information was compiled. One reported adverse event involved persistent mild leg pain following intraosseous insertion; no serious adverse events were reported.^{1,2}

Critique

PARAMEDIC-3 did not establish a survival advantage from starting with intraosseous access. It also did not establish equivalence. Recruitment stopped with 6,096 of 15,000 intended participants, leaving a wide interval around the adjusted survival estimate. Funding-related termination while investigators were blinded reduces concern about stopping opportunistically on a favourable result, but cannot restore lost precision. The neutral P value should be interpreted alongside the plausible range of benefit and harm.^{1,2}

The absolute scale is especially informative when survival is uncommon. The unadjusted interval allowed intraosseous-first treatment to reduce thirty-day survival by as much as 1.7 percentage points or improve it by 0.5 points. Either direction can matter in a high-volume emergency intervention. Adjusted estimates also remained uncertain. Stable accumulating results do not establish what another nine thousand participants would have shown, and no noninferiority margin was specified to support interchangeability.^{1,2}

The trial's strongest operational finding is the separation between insertion success and drug delivery. Intraosseous access substantially increased first-attempt success, yet median call-to-drug time was identical. Successful insertion is one step in a resuscitation sequence containing several simultaneous priorities. The data do not identify which subsequent steps offset the technical advantage. They do show that choosing the easier initial access route did not automatically accelerate the intended pharmacological intervention in these services.^{1,2}

This is a comparison of initial strategies, rather than exclusive use of one route throughout resuscitation. Nearly one-third of intravenous-first patients had intraosseous access as their first successful route. Some were protocol crossovers, while others received appropriate rescue after two failed attempts. Labelling all of them nonadherent would misdescribe the intervention. The intention-to-treat estimate answers whether services should begin with one strategy while retaining rescue options; it cannot isolate the effect of an intraosseous versus intravenous drug dose in otherwise identical patients.^{1,2}

The asymmetric crossover rate may have reduced the route contrast, but analyses restricted by actual treatment cannot remove this problem without introducing another. Difficulty obtaining access, prolonged resuscitation and clinician decisions can all be associated with prognosis. Selecting patients according to successful delivery after randomisation would sacrifice part of the protection provided by random allocation. Such analyses are supportive sensitivity checks, not superior causal estimates of route efficacy.^{1,2}

Adjustment for time to drug administration deserves particular attention because timing was a proposed mechanism of benefit. A variable influenced by allocation can lie on the causal pathway, so including it in the primary model complicates interpretation of the total strategy effect. The unadjusted result and post hoc models excluding drug timing or substituting response time remained neutral. That consistency is reassuring, although it does not make the modelling choices conceptually interchangeable or establish the absence of a timing-mediated effect.^{1,2}

The circulation outcomes should neither be ignored nor promoted above the primary endpoint. Both any and sustained return of circulation favoured intravenous-first treatment, with intervals excluding no difference. Their consistency gives the signal clinical credibility, but multiplicity, secondary status and the absence of a clear discharge or thirty-day survival effect limit certainty. These findings support caution about routine substitution; they do not prove that intraosseous drug delivery causes worse neurological survival.^{1,2}

Possible mechanisms include differences in drug transit or marrow perfusion during low-flow resuscitation. PARAMEDIC-3 did not measure circulating drug concentrations or randomise insertion site, so these remain hypotheses. Predominantly tibial use also means the overall result cannot determine whether humeral access is better or worse. Comparisons between patients receiving different sites would be vulnerable to clinician selection and differences in access difficulty.^{1,2}

Concealed allocation, a large multicentre cohort and near-complete primary ascertainment are substantial strengths. The primary endpoint is objective and resistant to the observer expectations that complicate open procedural studies. However, treatment and outcome assessment were unmasked. This matters more for neurological scoring, recognition of adverse events and discretionary cointerventions than for vital status. The published report did not characterise hospital management or cardiopulmonary resuscitation quality sufficiently to assess whether these differed.^{1,2}

Neurological recovery was clinically important but less complete than thirty-day survival. The discharge analysis lacked scores for 102 patients, compared with only 165 observed favourable outcomes. Missingness therefore deserves more attention than the excellent primary follow-up rate alone suggests. A modified Rankin score of three permits substantial assistance with daily activities and should not be described as intact neurological recovery. Longer follow-up was outside the initial report, and discharge outcomes were not final for everyone still hospitalised.^{1,2}

Safety reporting was sparse. One mild pain event and no reported serious complications do not demonstrate equal procedural safety. Most patients died, limiting assessment of pain and later limb problems, and routine emergency documentation may not capture every infiltration or displacement. The trial's more useful safety contribution is its patient-centred and circulation outcomes; rare device complications remain uncertain.^{1,2}

Applicability is strongest to adults with ongoing arrest who still required access when trial-trained paramedics became involved. More than a third of screened patients already had access and were excluded. Median drug administration occurred twenty-four minutes after the emergency call, and shockable rhythms were uncommon. The findings should not be automatically extended to much earlier treatment, children, in-hospital arrests or a policy mandating a different intraosseous site. They nevertheless address a common operational choice in established ambulance systems.^{1,2}

Body of Evidence

PARAMEDIC2 randomised 8,014 patients to adrenaline or placebo during out-of-hospital cardiac arrest. Thirty-day survival was 3.2% versus 2.4%, odds ratio 1.39 (95% confidence interval 1.06–1.82), but favourable neurological survival at discharge did not clearly improve. The study supports testing drug delivery against meaningful outcomes rather than assuming that more successful resuscitation processes necessarily produce better recovery.³

ALPS addressed another drug question, comparing amiodarone, lidocaine and placebo in shock-refractory ventricular fibrillation or pulseless ventricular tachycardia. In its 3,026-patient per-protocol population, discharge survival was 24.4%, 23.7% and 21.0%, respectively, without a statistically clear overall improvement from either active drug. Route was not the randomised intervention. It therefore informs the uncertainty surrounding resuscitation pharmacology, but cannot establish which vascular-access strategy is preferable.⁴

A smaller randomised access trial by Reades and colleagues compared tibial intraosseous, humeral intraosseous and peripheral intravenous insertion in 182 patients. First-attempt success was 91%, 51% and 43%, respectively. Its technical endpoint demonstrated that insertion site matters and that intraosseous access is not a single uniform procedure. It was not a definitive survival trial. PARAMEDIC-3 advances that earlier work by testing a service-level initial strategy with patient-centred follow-up.⁵

IVIO, published online alongside PARAMEDIC-3 in October 2024, provided a second substantial randomised comparison. Among 1,479 analysed Danish patients, sustained return of circulation occurred in 30% with initial intraosseous access and 29% with initial intravenous access, risk ratio 1.06 (0.90–1.24). Thirty-day and favourable neurological survival were also uncertain. Intraosseous access was more often successful within two attempts, reinforcing the distinction between technical reliability and demonstrated clinical benefit.⁶

The two large 2024 reports agree that an intraosseous-first policy did not deliver a clear patient-centred advantage. They do not give identical estimates for circulation outcomes: PARAMEDIC-3 favoured intravenous-first treatment, whereas IVIO's estimate was close to neutral. Different populations, systems and endpoint definitions matter; PARAMEDIC-3 required sustained circulation to hospital handover, while IVIO used a different duration-based definition. Their results should not be treated as if they were duplicate measurements of precisely the same endpoint.^{1,6,7}

Nielsen's accompanying editorial supported retaining intravenous access as the initial approach while recognising intraosseous access as an important alternative. It emphasised the value of these pragmatic randomised comparisons and the potential significance of PARAMEDIC-3's circulation signal. That interpretation is reasonable as a response to the absence of demonstrated superiority, but is not evidence of a proven intravenous survival benefit. Differences in drug timing between the two trials cannot by themselves establish a causal explanation for their results.⁷

The BMJ retracted VICTOR on 18 December 2025 following randomisation and allocation concerns. Cited in the 2024 literature, its results are excluded from this appraisal. This later correction does not add a new trial to the book's 2024 evidence period.⁸

The practical advance is consequently a better-defined limit to the argument for routine intraosseous-first treatment. Easier access alone is insufficient: the strategy must improve the timely delivery and effectiveness of care, and ultimately survival with acceptable function. PARAMEDIC-3 does not undermine rescue intraosseous access after intravenous failure, because that option was explicitly part of its comparator. It leaves uncertainty about earlier delivery and specific insertion sites, while providing no reason to assume that replacing an intravenous-first strategy improves thirty-day survival.^{1,2}

Should intraosseous access be the first strategy during adult out-of-hospital cardiac arrest?

PARAMEDIC-3 found no thirty-day survival benefit from an intraosseous-first strategy, despite greater first-attempt success. Secondary circulation outcomes favoured intravenous-first treatment, and the trial does not establish equivalence.

References

1. Couper K, Ji C, Deakin CD, Fothergill RT, Nolan JP, Long JB, et al.; PARAMEDIC-3 Collaborators. A Randomized Trial of Drug Route in Out-of-Hospital Cardiac Arrest. *N Engl J Med* 2025;392(4):336–348. Published online 31 October 2024.
2. PARAMEDIC-3 Collaborators. Supplementary appendix to: A Randomized Trial of Drug Route in Out-of-Hospital Cardiac Arrest. *N Engl J Med* 2024. doi:10.1056/NEJMoa2407780.
3. Perkins GD, Ji C, Deakin CD, Quinn T, Nolan JP, Scamperin C, et al.; PARAMEDIC2 Collaborators. A Randomized Trial of Epinephrine in Out-of-Hospital Cardiac Arrest. *N Engl J Med* 2018;379(8):711–721.

4. Kudenchuk PJ, Brown SP, Daya M, Rea T, Nichol G, Morrison LJ, et al.; Resuscitation Outcomes Consortium Investigators. Amiodarone, Lidocaine, or Placebo in Out-of-Hospital Cardiac Arrest. *N Engl J Med* 2016;374(18):1711–1722.
5. Reades R, Studnek JR, Vandeventer S, Garrett J. Intraosseous versus intravenous vascular access during out-of-hospital cardiac arrest: a randomized controlled trial. *Ann Emerg Med* 2011;58(6):509–516.
6. Vallentin MF, Granfeldt A, Klitgaard TL, Mikkelsen S, Folke F, Christensen HC, et al. Intraosseous or Intravenous Vascular Access for Out-of-Hospital Cardiac Arrest. *N Engl J Med* 2025;392(4):349–360. Published online 31 October 2024.
7. Nielsen N. The Way to a Patient's Heart — Vascular Access in Cardiac Arrest. *N Engl J Med* 2025;392(4):398–400. Published online 31 October 2024.
8. The BMJ. Retraction: Intraosseous versus intravenous vascular access in upper extremity among adults with out-of-hospital cardiac arrest: cluster randomised clinical trial (VICTOR trial). *BMJ* 2025;391:r2558. Published 18 December 2025. doi:10.1136/bmj.r2558.

Vallentin MF, Granfeldt A, Klitgaard TL, Mikkelsen S, Folke F, Christensen HC, et al. Intraosseous or Intravenous Vascular Access for Out-of-Hospital Cardiac Arrest. N Engl J Med 2025;392(4):349–360. Published online 31 October 2024.

Introduction

Intravenous cannulation provides a familiar route for resuscitation drugs, but finding a vein during cardiac arrest can be difficult. Intraosseous insertion may be more reliable, although successful placement does not necessarily establish equally rapid delivery into the central circulation. Clinicians must choose without interrupting other essential treatment.¹

Before IVIO, recommendations favouring initial intravenous access rested on limited comparative evidence. Observational studies often associated intraosseous administration with poorer outcomes, but the route could have been chosen because intravenous attempts had already failed or resuscitation was prolonged. This creates a timing-related selection problem: the group receiving rescue intraosseous access may have a worse prognosis before that route is used. Statistical adjustment cannot guarantee removal of such differences.¹

Randomising the initial access strategy offers a more direct test. It also separates two questions that are easily conflated: whether an approach establishes access more successfully, and whether it improves circulation or subsequent recovery. Return of spontaneous circulation is close to the intended drug mechanism and more frequent than long-term survival, making it a practical primary endpoint. It remains an intermediate outcome rather than the final purpose of resuscitation.¹

IVIO compared initial intraosseous and intravenous attempts across Denmark's emergency medical services. Within the intraosseous group, a second randomisation compared humeral and tibial insertion. The central hypothesis was that initial route would alter sustained return of spontaneous circulation; the broader question was whether any procedural advantage translated into survival with favourable neurological function.^{1,2}

Synopsis

IVIO was an investigator-initiated, parallel-group superiority trial conducted through emergency medical services in all five Danish regions. Recruitment began in Central Denmark on 1 March 2022, with other regions joining progressively, and ended on 5 May 2024. The Danish system generally sends an ambulance and a physician-staffed unit to arrests. Danish foundations and academic institutions funded the trial.^{1,2}

Adults aged eighteen years or older were eligible when vascular access was indicated during out-of-hospital cardiac arrest. Exclusions were suspected traumatic arrest, functioning access already in place and previous trial enrolment. Initially, emergency enrolment required oral consent from an independent physician; this requirement was waived after the trial's regulatory transition in July 2022. Written consent was subsequently sought from survivors who regained capacity or, for incapacitated survivors, from a relative and an independent physician.¹

The responsible clinician opened an opaque, sealed envelope to assign intraosseous or intravenous access equally. Intraosseous patients were further randomised equally to humeral or tibial insertion. The devices were EZ-IO for intraosseous access and Venflon Pro Safety for intravenous cannulation. Clinicians completed an online training module and received practical instruction if not already proficient with intraosseous insertion. Treating teams knew allocation; patients, surrogates and interview-based outcome assessors were not informed of it.¹

Allocation governed the first two attempts, each defined by skin penetration. After two failed attempts, further access was at the clinician's discretion. Clinical judgement determined whether access functioned. Times to successful access and adrenaline administration were recorded by clinicians. At two sites, when a CT scan was clinically indicated and the intraosseous device remained in place, the scan could be extended to include its insertion area. Imaging was not required in every patient.^{1,2}

The primary outcome was sustained return of spontaneous circulation, defined as a palpable pulse or other evidence of circulation without further chest compressions for at least twenty minutes. Key secondary outcomes were thirty-day survival and thirty-day survival with a modified Rankin Scale score of zero to three. Additional outcomes included ninety-day survival and function, survivor quality of life, procedural success and complications. Longer follow-up was ongoing when the initial report was published.¹

Assuming sustained circulation in 30% versus 38%, corresponding to a risk ratio of 1.27, a sample of 1,470 provided 90% power at a two-sided alpha of 0.05. An independent committee reviewed safety after approximately fifty, two hundred, four hundred and

eight hundred enrolments. No formal stopping boundaries for benefit, harm or futility were specified in the main report. Recruitment reached the planned analysis sample.¹

The modified intention-to-treat population included randomised patients who fulfilled eligibility criteria at allocation. Binary outcomes were presented as unadjusted risk ratios and absolute differences with 95% confidence intervals. Fisher's exact testing followed a hierarchy from the primary outcome to the two key secondary outcomes. Because the primary result was not significant, additional P values for those secondary outcomes were not calculated. Other outcomes and subgroup analyses were exploratory, with no multiplicity-adjusted intervals.¹

Sensitivity analyses adjusted for age, witnessed status, early cardiopulmonary resuscitation and initial rhythm using modified Poisson regression. Further analyses excluded patients with no access attempts, then those whose first two attempts departed from allocation. Six patients had no recorded rhythm analysis before randomisation and regained circulation at the next assessment; they were assigned a nonshockable rhythm for the adjusted analyses in a decision explicitly described as post hoc. This did not alter the unadjusted primary comparison.^{1,2}

Among 4,618 screened patients, 1,506 were randomised: 746 intraosseous and 760 intravenous. Twenty-seven erroneous randomisations were excluded, leaving 731 and 748 for analysis. Exclusions concerned eligibility at randomisation. Another 781 eligible patients were not randomised, including cases where enrolment was forgotten, envelopes were unavailable or clinicians prioritised other care.^{1,2}

Mean age was approximately sixty-nine years and 70% were men. Most arrests occurred at home, and 77% had an initial nonshockable rhythm. Bystander resuscitation was recorded in 84% versus 83% of arrests not witnessed by emergency personnel. Median first-unit response times were seven versus six minutes. Primary follow-up was complete. Two intraosseous and one intravenous patient lacked thirty- and ninety-day survival information.^{1,2}

Forty-four intraosseous patients, or 6%, and twenty-five intravenous patients, or 3%, never underwent an access attempt, usually because resuscitation ended or circulation returned early. A further sixteen versus sixty-four received an attempt discordant with their assigned route within the first two attempts. First-attempt success was 623/731, or 85%, versus 456/748, or 61%; success within two attempts was 669/731, or 92%, versus 595/748, or 80%. These percentages retain all eligible allocated patients.¹

Median time to successful access was fourteen minutes in both groups and median time to adrenaline was fifteen minutes in both. These clocks began at cardiac-arrest notification: the emergency call or arrest time, whichever was later. Access timing included only successful insertions; drug timing included only recipients. Adrenaline was administered to 615 patients, or 84%, versus 621, or 83%, with a median of two doses. Recorded compression fraction in an available subset was 84% versus 83%, with similar compression rates.^{1,2}

Sustained circulation occurred in 221/731 patients, or 30%, versus 214/748, or 29%, risk ratio 1.06 (95% confidence interval 0.90–1.24; P=0.49). The absolute difference was 1.6 percentage points (–3.0 to 6.3). At thirty days, 85/729, or 12%, versus 75/747, or 10%, were alive, risk ratio 1.16 (0.87–1.56). Favourable neurological survival occurred in 67/728, or 9%, versus 59/746, or 8%, risk ratio 1.16 (0.83–1.62). Adjusted and adherence-based analyses did not materially change the interpretation.^{1,2}

At ninety days, 82/729 versus 71/747 patients were alive, risk ratio 1.18 (0.88–1.60). Favourable neurological survival occurred in 75/727 versus 64/746, risk ratio 1.20 (0.88–1.65). Survivor quality-of-life measures showed no clear between-group difference. Extravasation during resuscitation was recorded in five patients per group. Among patients achieving sustained circulation or receiving extracorporeal resuscitation, ventricular tachyarrhythmia occurred in 37/232 versus 25/219; no compartment syndrome, osteomyelitis or traumatic bone fracture was reported.^{1,2}

The intraosseous site comparison included 361 humeral and 370 tibial allocations. Sustained circulation occurred in 108, or 30%, versus 113, or 31%, risk ratio 0.98 (0.79–1.22). Thirty-day favourable neurological survival was 25/359 versus 42/369, risk ratio 0.61 (0.38–0.98), an exploratory result. Clinician-assessed displacement was 16/333 versus 3/349 among those obtaining access. In the selected CT subset, catheter tips lay within marrow in 23/32 humeral devices and all 35 tibial devices.^{1,2}

Critique

IVIO provides a clear answer to its superiority hypothesis: initial intraosseous access did not significantly increase sustained return of spontaneous circulation. The uncertainty is nevertheless clinically meaningful. Its primary interval permits approximately three fewer to six more patients with sustained circulation per hundred treated. That range is narrower than would be obtained from a small feasibility study, but it does not establish that the routes are interchangeable.¹

The study reached its intended size, but was powered for an eight-percentage-point difference in an intermediate endpoint. Thirty-day survival was much less common and had wider relative uncertainty. Its risk-ratio interval ranged from 0.87 to 1.56; the absolute difference ranged from -1.6 to 4.8 percentage points. A neutral circulation result cannot be used to exclude worthwhile differences in survival or recovery. No noninferiority margin was tested.¹

The primary endpoint has a coherent mechanistic rationale but an important limit. Twenty minutes without compressions is a major improvement over no circulation, yet patients may subsequently rearrest or die from brain injury and other complications. The trial appropriately measured longer-term survival and function. Those outcomes deserve emphasis even though they lacked power for small differences; the numerically higher intraosseous survival estimates remain uncertain rather than evidence of a delayed benefit.¹

Hierarchical testing was handled conservatively. Once the primary test failed, the key secondary outcomes were not assigned confirmatory P values. The same restraint must apply to the site comparison and subgroup findings. The humeral-versus-tibial neurological interval narrowly excluded no difference at thirty days, but arose among numerous exploratory analyses in a smaller population. It is a signal for further investigation, not definitive proof that tibial placement improves neurological outcome.^{1,2}

The operational data are particularly instructive. Intraosseous access succeeded more often within two attempts, but median access and adrenaline times were unchanged. These timing comparisons condition on eventually obtaining access or receiving the drug, excluding patients in whom the process never succeeded. They therefore cannot describe an unconditional time-to-treatment effect in everyone randomised. The supplement's curves explicitly examine death and return of circulation as competing events, providing a more complete picture than successful-procedure medians alone.^{1,2}

Some participants never received an attempt after allocation, more often in the intraosseous arm. Retaining them preserves the initial strategy comparison. Their subsequent clinical course determined whether access remained necessary, so simply removing them can select different prognostic groups. The adjusted sensitivity estimates were reassuringly consistent, but exclusions based on post-randomisation events cannot produce a perfectly unbiased estimate of treatment received.^{1,2}

Likewise, rescue and crossover have different meanings. Clinicians were allowed to change route after two failed attempts; this was part of the strategy. Switching earlier

was a departure, more common after intravenous allocation. The resulting estimate includes the practical availability of rescue access. It should not be interpreted as a pure pharmacokinetic comparison or as evidence against intraosseous use when initial intravenous attempts fail.^{1,2}

The additional site randomisation is a strength because site choice was not simply left to clinician preference. However, adherence differed: forty-five humeral and eleven tibial patients had an early attempt discordant with site allocation, beyond those receiving no attempt. The site comparison was not powered for modest patient-centred effects. Its circulation interval allows meaningful relative differences in either direction, despite similar point estimates.^{1,2}

CT findings require still greater caution. Only sixty-seven patients underwent the expanded scan, and analyses grouped them by actual successful route rather than original allocation. Imaging occurred roughly two hours after arrest notification. Nine humeral tips were outside marrow, but six of these patients had the arm positioned above the head during scanning despite instructions to keep it beside the body. The images cannot establish whether displacement occurred during insertion, resuscitation, transport or positioning. They raise a device-management concern without explaining the main clinical result.²

Concealed envelopes, complete primary ascertainment and blinded follow-up interviewers support validity. Clinicians were unmasked and judged access success and circulation at the scene. The few eligibility-based exclusions are less concerning than exclusions because treatment failed.^{1,2}

The compression data and post-arrest treatment records improve understanding of cointerventions. Compression measurements were available for only 574 patients from two regions, and hospital treatment summaries concerned twenty-four-hour survivors. Neither subset can demonstrate equivalent care throughout the entire randomised population. Nevertheless, broadly similar recorded compression performance reduces concern that one access procedure systematically displaced resuscitation in the observed cases.²

Reported complications were uncommon, but ascertainment windows and denominators matter. Extravasation was assessed during resuscitation, whereas arrhythmias and later local complications were evaluated after sustained circulation or extracorporeal support. Comparisons among survivors can be influenced by who survives to be assessed. Absence of a reported rare complication does not prove absence of risk, particularly when most participants died early.²

Generalisability is strongest to a trained, physician-supported Danish system with high bystander resuscitation rates. The regions joined at different times, and Central Denmark supplied more than half the analysed sample. Hundreds of otherwise eligible patients were not randomised for operational or clinical reasons. Other services may have different intravenous skill, device familiarity, staffing and response times. These differences may affect implementation.^{1,2}

Body of Evidence

Earlier evidence addressed different parts of the pathway. Reades and colleagues' 182-patient randomised trial focused on technical access success and favoured tibial intraosseous insertion. It did not establish a survival advantage. IVIO adds larger clinical comparisons of both sites.³

The ALPS route analysis examined 3,019 patients with known access who had been randomised to an antiarrhythmic drug or placebo. Drug-associated survival estimates appeared more favourable among intravenous recipients, but formal interactions with route were not significant. Access itself was not randomised. Differences between route groups and limited interaction power prevented that analysis from establishing a preferred access strategy; it supplied a reason to conduct a direct trial such as IVIO.⁴

The Danish COCA trial tested calcium rather than access route. Among 391 analysed patients, sustained circulation occurred in 19% with calcium and 27% with saline, risk ratio 0.72 (95% confidence interval 0.49–1.03). Recruitment stopped because of concern about harm. Drugs could be given intravenously or intraosseously, but route was not randomised. This distinction matters: physiological plausibility, route availability and the efficacy of the administered drug are separate questions.⁵

PARAMEDIC-3 supplied the largest contemporary direct comparison. In its 6,082-patient analysis population, thirty-day survival was 4.5% with intraosseous-first and 5.1% with intravenous-first treatment, adjusted odds ratio 0.94 (0.68–1.32). Any and sustained return of circulation favoured intravenous-first treatment as secondary outcomes. Recruitment ended early for funding reasons. Together, the trials show no demonstrated clinical superiority from routinely beginning with intraosseous access, while leaving uncertainty about the size and direction of smaller effects.⁶

Differences between IVIO and PARAMEDIC-3 deserve explanation rather than a contest between P values. The UK trial primarily used tibial intraosseous access, whereas IVIO randomised between two sites. IVIO used twenty minutes without compressions for its primary outcome; PARAMEDIC-3 defined sustained circulation through hospital

handover. Their drug-timing clocks were not identical, and baseline prognosis and service organisation differed. These features limit simple numerical comparisons and prevent attributing the different circulation estimates to any one factor.^{1,2,6}

Nielsen's accompanying editorial favoured retaining intravenous access as the initial strategy, with intraosseous access available when needed. It recognised the high technical success of intraosseous insertion while giving weight to the absence of patient-centred benefit and PARAMEDIC-3's circulation signal. This is a defensible practical interpretation, provided it is not converted into a claim that intravenous access has conclusively improved survival in both trials.⁷

The 2024 discussion also cited VICTOR, a cluster-randomised access trial. The BMJ retracted that report on 18 December 2025 following concerns about randomisation and allocation. This later integrity update is disclosed separately from the historical evidence period; VICTOR's results are not used to support the chapter's conclusions.⁸

IVIO demonstrates that greater access success need not improve drug timing or recovery. Its site findings need patient-centred confirmation. The trial supports reliable rescue options, but neither establishes equal efficacy nor justifies declaring a site universally unsafe from a small CT subset.^{1,2}

Does initial intraosseous access improve resuscitation outcomes compared with intravenous access?

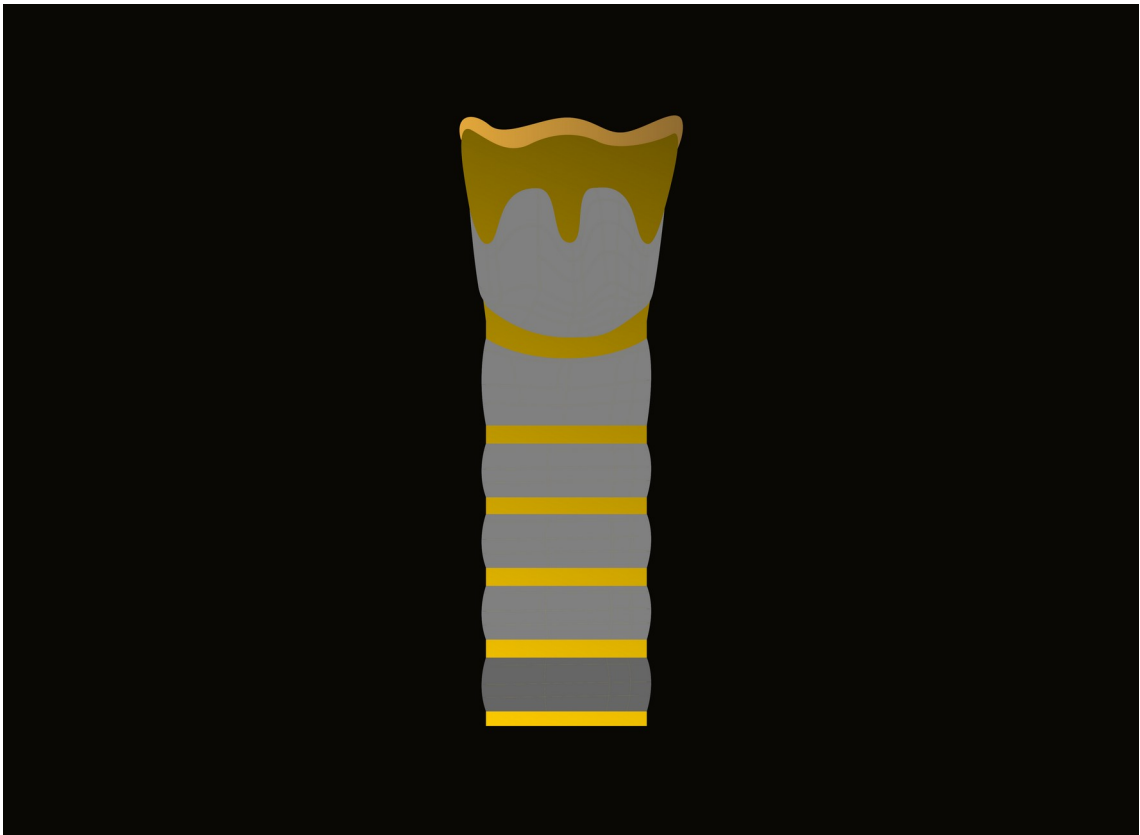
IVIO found no improvement in sustained return of circulation despite greater access success. Survival and neurological outcomes remained uncertain, and neither the main comparison nor the smaller site comparison establishes equivalence.

References

1. Vallentin MF, Granfeldt A, Klitgaard TL, Mikkelsen S, Folke F, Christensen HC, et al. Intraosseous or Intravenous Vascular Access for Out-of-Hospital Cardiac Arrest. *N Engl J Med* 2025;392(4):349–360. Published online 31 October 2024.
2. Vallentin MF, Granfeldt A, Klitgaard TL, Mikkelsen S, Folke F, Christensen HC, et al. Supplementary appendix to: Intraosseous or Intravenous Vascular Access for Out-of-Hospital Cardiac Arrest. *N Engl J Med* 2024. doi:10.1056/NEJMoa2407616.

3. Reades R, Studnek JR, Vandeventer S, Garrett J. Intraosseous versus intravenous vascular access during out-of-hospital cardiac arrest: a randomized controlled trial. *Ann Emerg Med* 2011;58(6):509–516.
4. Daya MR, Leroux BG, Dorian P, Rea TD, Newgard CD, Morrison LJ, et al. Survival After Intravenous Versus Intraosseous Amiodarone, Lidocaine, or Placebo in Out-of-Hospital Shock-Refractory Cardiac Arrest. *Circulation* 2020;141(3):188–198.
5. Vallentin MF, Granfeldt A, Meilandt C, Povlsen AL, Sindberg B, Holmberg MJ, et al. Effect of Intravenous or Intraosseous Calcium vs Saline on Return of Spontaneous Circulation in Adults With Out-of-Hospital Cardiac Arrest: A Randomized Clinical Trial. *JAMA* 2021;326(22):2268–2276.
6. Couper K, Ji C, Deakin CD, Fothergill RT, Nolan JP, Long JB, et al.; PARAMEDIC-3 Collaborators. A Randomized Trial of Drug Route in Out-of-Hospital Cardiac Arrest. *N Engl J Med* 2025;392(4):336–348. Published online 31 October 2024.
7. Nielsen N. The Way to a Patient's Heart — Vascular Access in Cardiac Arrest. *N Engl J Med* 2025;392(4):398–400. Published online 31 October 2024.
8. The BMJ. Retraction: Intraosseous versus intravenous vascular access in upper extremity among adults with out-of-hospital cardiac arrest: cluster randomised clinical trial (VICTOR trial). *BMJ* 2025;391:r2558. Published 18 December 2025. doi:10.1136/bmj.r2558.

Possible Reporting Inconsistencies. Table S22 gives the humeral-versus-tibial risk difference for sustained return of circulation as –0.6 percentage points with a 95% confidence interval of –0.7 to 6.0. The lower limit appears misprinted and understates uncertainty on the side favouring tibial access. The chapter uses the published risk ratio and its interval, 0.98 (0.79–1.22).



Airway Trials

PREOXI

Gibbs KW, Semler MW, Driver BE, Seitz KP, Stempek SB, Taylor C, et al.; PREOXI Investigators and the Pragmatic Critical Care Research Group. Noninvasive Ventilation for Preoxygenation during Emergency Intubation. N Engl J Med 2024;390(23):2165–2177.

Introduction

Emergency intubation interrupts breathing at a time when oxygen reserves may already be depleted. Preoxygenation aims to increase the oxygen available during induction and laryngoscopy, but a high oxygen flow through a face mask cannot necessarily overcome severe shunt, loss of lung volume or inadequate ventilation. A satisfactory starting saturation also does not guarantee protection throughout the procedure.¹

Noninvasive ventilation offers a different approach. A close-fitting interface can provide a high inspired oxygen concentration, while positive end-expiratory pressure and inspiratory assistance may support lung recruitment and ventilation. These physiological advantages must be balanced against practical difficulties, including mask intolerance, preparation time and concern about gastric insufflation or aspiration. The question is particularly relevant when a deteriorating patient needs intubation but still has time for several minutes of preparation.¹

Earlier trials suggested that positive pressure could improve peri-intubation oxygenation, but they were smaller or addressed narrower populations and different stages of the procedure. Choosing a preoxygenation device also influences what support is easily continued after induction. Consequently, an apparently simple comparison of devices may test a broader oxygenation strategy extending into the interval before laryngoscopy.¹

PREOXI compared noninvasive ventilation with an oxygen mask across emergency departments and intensive care units. Its primary question was whether the noninvasive ventilation strategy reduced saturation below 85% during intubation. It also examined severe desaturation, haemodynamic complications and aspiration. The distinction between preventing a dangerous physiological event and demonstrating better survival remains central to interpreting its findings.¹

Synopsis

PREOXI was a pragmatic, unblinded, parallel-group randomised trial conducted in seven emergency departments and seventeen intensive care units at fifteen US medical centres. Recruitment ran from 10 March 2022 to 14 October 2023; the report was first published online on 13 June 2024. The US Department of Defense funded the trial without involvement in its design, conduct, analysis or reporting. Institutional review board approval included a waiver of informed consent, with information about participation subsequently provided to patients or their representatives.^{1,2}

Adults undergoing intubation with sedation and a laryngoscope were eligible. Exclusions included pregnancy, imprisonment, existing positive-pressure ventilation, apnoea or hypopnoea, urgency preventing trial procedures, and a clinician's judgement that either technique was required or contraindicated. Among 4,567 screened patients, 1,301 were randomised: 645 to noninvasive ventilation and 656 to an oxygen mask. Common exclusions included insufficient preparation time and existing positive-pressure support. Active vomiting or bleeding that prevented noninvasive ventilation and perceived high aspiration risk also restricted enrolment.^{1,2}

Randomisation used variable block sizes stratified by site, with allocation concealed in sequentially numbered opaque envelopes. Operators and research personnel knew the assignment. An observer who was not performing the intubation collected procedural observations, including the lowest reliable pulse oximetry value. Operators reported aspiration immediately afterwards. Radiologists interpreting subsequent chest images were unaware of treatment allocation.^{1,2}

The noninvasive ventilation strategy used a tight-fitting mask attached to a conventional ventilator or a dedicated noninvasive device. Recommended settings were an inspired oxygen fraction of 1.0, expiratory pressure of at least 5 cm H₂O, inspiratory pressure of at least 10 cm H₂O and a respiratory rate of at least ten per minute. The inspiratory pressure refers to the total inspiratory level, not an additional 10 cm H₂O above expiratory pressure. Preoxygenation for at least three minutes was recommended whenever feasible, with continuation after induction until laryngoscopy encouraged.^{1,2}

Controls received a non-rebreather or bag-mask device without manual ventilation during preoxygenation, using the highest available oxygen flow of at least 15 litres per minute. Bag-mask ventilation after induction was permitted in both groups. Standard or high-flow nasal oxygen could be used during preoxygenation or laryngoscopy, and clinicians selected other procedural treatments. The comparison therefore allowed usual clinical decisions around a defined initial oxygenation strategy.^{1,2}

The primary outcome was oxygen saturation below 85% between induction and two minutes after intubation. The secondary outcome was the lowest saturation during this interval. Exploratory outcomes included more severe desaturation, cardiovascular collapse, aspiration, new chest imaging abnormalities, ventilator-free days and death before hospital discharge or day twenty-eight. Cardiovascular collapse comprised systolic pressure below 65 mm Hg, new or increased vasopressor administration, or cardiac arrest.¹

A sample of 1,264 gave 85% power to detect a six-percentage-point reduction from an assumed 17% control event rate, using a two-sided significance level of 0.05. Allowance for missing data increased the target to 1,300. An independent monitoring committee reviewed the trial after 650 participants against a stringent interim efficacy boundary and recommended continuation without modification. The primary analysis compared patients according to allocation using available outcome data; additional models adjusted for site and selected clinical characteristics.^{1,2}

Median age was sixty-one years, approximately 60% were men, and median APACHE II score was seventeen. Nearly half had acute hypoxaemic respiratory failure. Approximately 73% were intubated in intensive care, 27% were already receiving vasopressors, and almost one quarter had received high-flow nasal oxygen before enrolment. Most procedures were performed by residents or fellows. Altered mental status was common, so the population was not confined to patients who were alert and readily cooperative.^{1,2}

Assigned treatment was delivered to 616 patients in the noninvasive ventilation group and 648 controls, corresponding to 95.5% and 98.8%. Actual median inspiratory and expiratory pressures were 10 and 5 cm H₂O. About 95% or more in each group received at least three minutes of preoxygenation. Positive-pressure ventilation after induction was substantially more frequent with the noninvasive strategy, at 88.1% versus 31.1%. Videolaryngoscopy, neuromuscular blockade and first-attempt intubation success were broadly similar.^{1,2}

The primary outcome was available for 624 and 637 patients. Desaturation below 85% occurred in 57 patients receiving noninvasive ventilation, 9.1%, and 118 controls, 18.5%. The absolute risk difference was -9.4 percentage points, with a 95% confidence interval of -13.2 to -5.6, $P < 0.001$. Adjustment gave a similar conclusion. Median lowest saturation was 99% versus 97%, a difference of two percentage points, with a 95% confidence interval of one to three.^{1,2}

Saturation below 80% occurred in 6.2% versus 13.2%, an absolute difference of –6.9 percentage points, while saturation below 70% occurred in 2.4% versus 5.7%, a difference of –3.2 percentage points. Oxygenation already differed at induction: saturation was 95% or lower in 8.3% versus 17.4%. These findings suggest that the intervention improved oxygenation before the subsequent period of apnoea, although they do not isolate the contribution of each treatment phase.¹

Cardiovascular collapse occurred in 17.5% versus 19.4%, with an absolute difference of –1.8 percentage points and a confidence interval from –6.1 to 2.4. Cardiac arrest occurred in one versus seven patients. Operator-reported aspiration occurred in six versus nine patients, 0.9% versus 1.4%, with a difference of –0.4 percentage points and a confidence interval from –1.6 to 0.7. New infiltrates occurred in 28.3% versus 29.8% of those with follow-up imaging. Two aspiration events in the noninvasive ventilation group were separately reported as serious adverse events; one was fatal.^{1,2}

Death before hospital discharge or day twenty-eight occurred in 209 patients, 32.4%, versus 217, 33.1%. The absolute difference was –0.7 percentage points, with a confidence interval from –5.8 to 4.4. Differences in ventilator-free and intensive care-free days were also uncertain. The trial established a reduction in peri-intubation hypoxaemia, without demonstrating an improvement in subsequent survival.¹

Critique

PREOXI provides a persuasive answer to its primary question. Recruitment reached the planned sample, allocation was concealed, adherence was high, and the effect was both statistically clear and clinically substantial. The finding was supported by lower saturation thresholds, the continuous saturation outcome and adjusted analyses. The observed absolute reduction corresponds to roughly eleven patients treated to prevent one episode below 85%, calculated from the published risk difference. This is a number needed to prevent desaturation, not a number needed to prevent death.^{1,2}

The multicentre pragmatic design is another strength. Equipment and operators reflected ordinary participating-unit practice rather than a single highly specialised team. Nevertheless, only about 28% of screened patients were randomised. The results are most directly applicable when clinicians consider either method acceptable and have time to prepare. They cannot establish the best strategy for an immediately crashing patient, someone already dependent on positive-pressure support, or someone whose vomiting or airway bleeding makes a tight mask unsuitable.^{1,2}

The most important interpretive limitation concerns the treatment contrast. Noninvasive ventilation was used before induction and commonly continued afterwards, whereas fewer controls received positive pressure after induction. Randomisation therefore estimates the effect of the whole strategy as implemented. It cannot divide the benefit into separate contributions from preoxygenation, recruitment and ventilation during the interval before laryngoscopy. Describing the trial as an isolated test of three minutes of preoxygenation would understate the actual difference between groups.^{1,2}

This issue was raised in the accompanying correspondence. Li and colleagues questioned the imbalance in post-induction ventilation and its implications for interpreting the control group. In reply, the investigators emphasised the oxygenation difference already present at induction and analyses across sites with differing use of bag-mask ventilation. The former directly supports a pre-induction physiological effect. The latter remains a non-randomised comparison and cannot prove that post-induction ventilation contributed nothing. Both observations can be true: preoxygenation improved, and continuing ventilation may have added benefit.^{3,4}

Blinding was impractical, and independent observation should not be mistaken for masked observation. Pulse oximetry is comparatively objective, but waveform selection, recording and rescue interventions still involve judgement. Forty patients lacked the primary outcome, mainly because a reliable waveform was unavailable. Missingness was numerically balanced, yet poor perfusion can link missing saturation measurements to the most serious complications. A complete-case analysis consequently has a different vulnerability from merely losing routine follow-up in otherwise stable patients.^{1,2}

The large effect is unlikely to be explained solely by those missing values. Even assigning hypoxaemia to all twenty-one missing patients in the intervention group and none of the nineteen missing controls would leave event proportions of approximately 12.1% and 18.0%. This simple, deliberately adverse calculation is not a prespecified trial analysis, but it shows that this particular missing-data scenario would not reverse the direction of the result. It does not resolve possible measurement errors among recorded observations.^{1,2}

The unadjusted primary comparison also deserves priority over more complicated sensitivity models. The supplement explains that acute hypoxaemic respiratory failure was recorded as an indication for intubation and was not strictly a pretreatment measurement. Including it in adjustment or subgroup analyses therefore requires caution. Reassurance comes from the strong unadjusted result and the similar model accounting for site alone, rather than from assuming that every adjustment variable was immune to treatment-related classification.²

Safety findings are encouraging but less definitive. Aspiration was uncommon and identified principally through operator reporting. Clinicians could exclude patients considered at especially high risk, and nearly three hundred patients lacked the follow-up chest imaging used for some pulmonary outcomes. A new infiltrate is also not specific for aspiration. The confidence interval around aspiration permits either a reduction or an increase, and the two serious aspiration reports prevent a claim that no important adverse events occurred. Similar observed rates do not establish equal safety in excluded high-risk patients.^{1,2}

The one-versus-seven cardiac arrest comparison is clinically notable but rests on eight events within a wider exploratory analysis. It should not be promoted to proof of a survival benefit. Conversely, the absence of a mortality difference does not negate the established physiological benefit: the trial was not powered for modest mortality effects, and hospital mortality reflects many processes beyond intubation. Its follow-up definition also censored observation at discharge or day twenty-eight, rather than establishing complete twenty-eight-day vital status for everyone after discharge.¹

Implementation requires more than selecting a device. Staff must establish a seal, select appropriate pressures, coordinate induction and remove the interface for laryngoscopy. The investigators noted that a conventional ventilator and circuit could sometimes be used for both preoxygenation and subsequent invasive ventilation. That addresses some practical concerns raised in correspondence, but it is not a trial-based economic or environmental evaluation. Local equipment and staffing will determine the additional work required.^{3,4}

Finally, high-flow nasal oxygen was permitted but not randomised, and its use differed between groups. Prior use in almost one quarter of participants generated understandable concern about changing an established support modality before intubation. PREOXI does not settle whether noninvasive ventilation is superior to a consistently applied high-flow strategy, whether the two should be combined, or which patients benefit most from either approach. Those are distinct comparisons.^{1,4,5}

Body of Evidence

The rationale for noninvasive preoxygenation predates PREOXI. In a two-centre trial of fifty-three hypoxaemic patients, Baillard and colleagues compared three minutes of noninvasive ventilation with usual mask preoxygenation. Saturation below 80% occurred in two of twenty-seven versus twelve of twenty-six patients. This small study established a substantial physiological signal, but its size could not reliably settle uncommon harms or patient-centred outcomes. PREOXI strengthens that earlier signal in a much larger and more varied population.⁶

A subsequent six-centre trial by Baillard and colleagues enrolled 201 hypoxaemic patients and examined maximum SOFA score during the first seven days after intubation. Median scores were nine with noninvasive ventilation and ten with mask preoxygenation, $P=0.65$. This different primary endpoint illustrates why improved oxygenation cannot automatically be translated into reduced organ failure. Together, the trials support a distinction between an effective procedural intervention and an unproven effect on the subsequent course of critical illness.⁷

The PreVent trial addressed the interval after induction. Among 401 critically ill adults, bag-mask ventilation increased median lowest saturation from 93% to 96%. Saturation below 80% occurred in 10.9% versus 22.8%, relative risk 0.48, with a 95% confidence interval of 0.30 to 0.77. Aspiration was not significantly different. Its relevance to PREOXI is mechanistic: ventilation during this interval can itself affect oxygenation. It does not establish which device is best before induction or quantify the contribution of this phase within PREOXI.⁸

FLORALI-2 tested another comparator: noninvasive ventilation against high-flow nasal oxygen with apnoeic oxygenation in acute hypoxaemic respiratory failure. In its 313-patient analysis, severe hypoxaemia occurred in 23% versus 27%, an absolute difference of -4.2 percentage points, with a confidence interval from -13.7 to 5.5 , $P=0.39$. This was not an overall positive superiority result. The different comparator and population mean that PREOXI cannot simply be used to overturn that uncertainty or claim superiority over every alternative oxygenation technique.⁹

Taken together, these studies locate PREOXI's contribution precisely. It provides strong evidence that a strategy using noninvasive ventilation around induction reduces hypoxaemia compared with oxygen-mask preoxygenation and discretionary subsequent ventilation. The evidence is especially useful because it concerns a familiar clinical decision with an immediate, measurable consequence. Its strength lies in the consistency and magnitude of the oxygenation effect, rather than in an established reduction in organ failure, neurological injury or mortality.¹

For practice, noninvasive ventilation is a well-supported option when the patient can receive it safely and preparation time is available. Applying the findings means reproducing the relevant strategy, including attention to ventilation between induction and laryngoscopy, rather than simply placing a mask briefly before giving drugs. Aspiration risk, tolerance and urgency remain reasons for individual judgement. Further comparisons should clarify high-flow alternatives, combinations of support and outcomes beyond the brief procedural window.^{1,2}

Does noninvasive ventilation reduce hypoxaemia during emergency intubation?

Yes. PREOXI reduced saturation below 85% from 18.5% to 9.1% with a strategy extending around induction, without establishing a survival benefit.

References

1. Gibbs KW, Semler MW, Driver BE, Seitz KP, Stempek SB, Taylor C, et al.; PREOXI Investigators and the Pragmatic Critical Care Research Group. Noninvasive Ventilation for Preoxygenation during Emergency Intubation. *N Engl J Med* 2024;390(23):2165–2177.
2. Gibbs KW, Semler MW, Driver BE, Seitz KP, Stempek SB, Taylor C, et al. Supplementary appendix to: Noninvasive Ventilation for Preoxygenation during Emergency Intubation. *N Engl J Med* 2024. doi:10.1056/NEJMoa2313680.
3. Li J, Scott JB, Branson R. Noninvasive Ventilation for Preoxygenation during Emergency Intubation [correspondence]. *N Engl J Med* 2024;391(11):1068–1069.
4. Gibbs KW, Semler MW, Casey JD. Noninvasive Ventilation for Preoxygenation during Emergency Intubation. Reply. *N Engl J Med* 2024;391(11):1069–1070.
5. Jacobs FM. Noninvasive Ventilation for Preoxygenation during Emergency Intubation [correspondence]. *N Engl J Med* 2024;391(11):1069.
6. Baillard C, Fosse JP, Sebbane M, Chanques G, Vincent F, Courouble P, et al. Noninvasive ventilation improves preoxygenation before intubation of hypoxic patients. *Am J Respir Crit Care Med* 2006;174(2):171–177.

7. Baillard C, Prat G, Jung B, Futier E, Lefrant JY, Vincent F, et al. Effect of preoxygenation using non-invasive ventilation before intubation on subsequent organ failures in hypoxaemic patients: a randomised clinical trial. *Br J Anaesth* 2018;120(2):361–367.
8. Casey JD, Janz DR, Russell DW, Vonderhaar DJ, Joffe AM, Dischert KM, et al.; PreVent Investigators and the Pragmatic Critical Care Research Group. Bag-Mask Ventilation during Tracheal Intubation of Critically Ill Adults. *N Engl J Med* 2019;380(9):811–821.
9. Frat JP, Ricard JD, Quenot JP, Pichon N, Demoule A, Forel JM, et al.; FLORALI-2 study group and REVA network. Non-invasive ventilation versus high-flow nasal cannula oxygen therapy with apnoeic oxygenation for preoxygenation before intubation of patients with acute hypoxaemic respiratory failure: a randomised, multicentre, open-label trial. *Lancet Respir Med* 2019;7(4):303–312.



Respiratory Trials

RENOVATE

RENOVATE Investigators and the BRICNet Authors. High-Flow Nasal Oxygen vs Noninvasive Ventilation in Patients With Acute Respiratory Failure: The RENOvATE Randomized Clinical Trial. JAMA 2025;333(10):875–890. Published online 10 December 2024.

Introduction

Noninvasive respiratory support can improve gas exchange while avoiding the burdens of tracheal intubation. High-flow nasal oxygen is often easier to tolerate than a tightly fitted ventilation mask and allows patients to communicate, eat and drink more readily. Noninvasive ventilation provides more direct assistance to breathing and controlled positive airway pressure, but discomfort, leakage and interrupted treatment can limit its delivery.¹

The best choice may depend on why respiratory failure has developed. Acute hypoxaemic lung disease, ventilatory failure during a COPD exacerbation and cardiogenic pulmonary oedema involve different physiological problems. Immunosuppression and COVID-19 further alter prognosis and the evidence informing treatment. A device that performs satisfactorily in one group cannot be assumed to provide equivalent support in all the others.¹

RENOVATE approached these related questions within one adaptive trial. It compared an initial high-flow strategy with face-mask noninvasive ventilation across distinct clinical groups, using a Bayesian model that could share information between them. This design promised greater efficiency than conducting a separate large trial for every diagnosis, but its conclusions depend partly on whether borrowing across groups is clinically and statistically appropriate.^{1,2}

The primary objective was noninferiority for intubation or death within seven days. Interpretation requires attention to the permitted margin, the direct evidence within each group and the rescue treatments patients received.^{1,2}

Synopsis

RENOVATE was an adaptive, multicentre, open-label randomised trial at thirty-three Brazilian hospitals. Adults were recruited from intensive care units, emergency departments and wards between November 2019 and November 2023. Final follow-up occurred in April 2024, and publication followed on 10 December 2024. The Brazilian Ministry of Health supported the study; Fisher & Paykel supplied high-flow equipment and disposables. The Hcor Research Institute coordinated the trial, with specialised statistical support from Berry Consultants.¹

Patients entered one of five final groups: nonimmunocompromised hypoxaemic respiratory failure, immunocompromised hypoxaemic respiratory failure, COPD exacerbation with respiratory acidosis, acute cardiogenic pulmonary oedema, or hypoxaemic COVID-19. Hypoxaemic eligibility generally required room-air saturation below 90% or arterial oxygen tension below 60 mm Hg, together with respiratory effort or tachypnoea. COPD eligibility required suspected or established COPD, respiratory distress, pH below 7.35 and arterial carbon dioxide tension above 45 mm Hg. Cardiogenic pulmonary oedema had separate clinical and oxygenation criteria.^{1,2}

Important exclusions included an immediate need for intubation, severe haemodynamic instability, markedly impaired consciousness, contraindications to noninvasive ventilation and an existing do-not-intubate order. Patients with cardiogenic pulmonary oedema who had already received noninvasive ventilation were excluded. Other eligible groups could have received a limited period of noninvasive support before randomisation. Thus, the trial concerned patients in whom continued noninvasive management remained a reasonable option.^{1,2}

An online system concealed allocation, using permuted blocks stratified by site and respiratory failure category. Patients and clinicians could not be blinded. Consent was sought after emergency enrolment. Of 1,800 randomised patients, thirty-four were excluded because they died before consent could be obtained and permission to use their data was unavailable. The analysis included 1,766 patients, with 883 in each treatment arm. Primary outcome data were complete within this consent-based population.¹

High-flow nasal oxygen began at 30 litres per minute for COPD and 45 litres per minute for other groups, increasing towards 60 litres per minute or the highest tolerated flow. Inspired oxygen was titrated to saturation of 88–92% in COPD and 92–98% otherwise. Noninvasive ventilation was delivered through a face mask, using inspiratory and expiratory pressures adjusted to clinical response and a target tidal volume of six to nine

millilitres per kilogram of ideal body weight. Continuous treatment during the first day was encouraged; low-flow oxygen could be used during ventilation breaks.^{1,2}

The high-flow strategy allowed rescue noninvasive ventilation for deteriorating COPD and cardiogenic pulmonary oedema patients. The published rescue pathway specified reassessment and up to one hour of rescue ventilation before intubation if failure criteria persisted. Intubation criteria addressed refractory gas-exchange abnormalities, worsening acidosis, instability, impaired consciousness, secretion problems and intolerance, although the attending clinician retained the final decision.^{1,2}

The primary outcome combined intubation and death within seven days. Death was added to the original intubation-only endpoint in April 2021. Secondary outcomes included twenty-eight- and ninety-day mortality and days free from invasive ventilation or intensive care by day twenty-eight; patients who died received a value of minus one for free-day outcomes. Comfort, length of stay and subsequent treatment limitations were also assessed.^{1,2}

COVID-19 was not anticipated when the trial began. Patients with COVID-19 were initially enrolled within the existing categories, then separated into a fifth analysis group following a decision by the steering committee while blinded to outcome data. The amendment was documented in June 2023. The model accounted for changes in COVID-19 outcome risk over time, and allowed adaptive stopping of individual groups for noninferiority, superiority or futility.^{1,2}

The noninferiority margin was an odds ratio of 1.55 for high-flow therapy versus noninvasive ventilation. At an assumed 30% control event rate, this corresponds approximately to allowing an increase to 40%. The final posterior probability threshold was 0.992. Superiority could be tested after noninferiority, using the same probability threshold for an odds ratio below one. Simulations informed stopping rules and error control, with up to six interim analyses and a maximum sample of two thousand.^{1,2}

The primary Bayesian hierarchical model borrowed information across diagnostic groups, drawing treatment estimates towards a common pattern to a degree determined by the model and observed data. It also incorporated prior distributions for control event rates. Analyses using weaker priors and the per-protocol population were accompanied by a post hoc analysis without cross-group borrowing. These analyses address different assumptions and should be considered together.^{1,2}

Mean age was approximately sixty-four years and 40% were women. There were 485 nonimmunocompromised hypoxaemic patients, fifty immunocompromised patients,

seventy-seven with COPD, 272 with cardiogenic pulmonary oedema and 882 with COVID-19. Assigned treatment was received by 93.3% of high-flow patients and 91.5% of ventilation patients. High-flow treatment was generally delivered for longer each day. Eight of thirty-five high-flow COPD patients, 22.9%, crossed to noninvasive ventilation, compared with less than 10% in each other high-flow group.^{1,2}

Overall, intubation or death occurred in 344 of 883 high-flow patients and 336 of 883 controls. The prespecified model met noninferiority criteria in four groups. Among nonimmunocompromised hypoxaemic patients, events occurred in 81 of 249 versus 78 of 236, 32.5% versus 33.1%; the modelled odds ratio was 1.02, with a 95% credible interval of 0.81–1.26 and noninferiority probability 0.999. In COVID-19, events occurred in 223 of 435 versus 210 of 447, 51.3% versus 47.0%; the odds ratio was 1.13, with an interval of 0.94–1.38 and probability 0.997.¹

In COPD, events occurred in ten of thirty-five versus eleven of forty-two, 28.6% versus 26.2%. The modelled odds ratio was 1.05, with an interval of 0.79–1.36; the more precise supplementary noninferiority probability was 0.9924. In cardiogenic pulmonary oedema, fourteen of 136 versus twenty-nine of 136 patients had events, 10.3% versus 21.3%, but borrowing produced an odds ratio of 0.97, with an interval of 0.73–1.23 and probability 0.997. No group met the primary model's superiority criterion.^{1,2}

The immunocompromised group stopped for futility at the first interim analysis. Its final event proportions were sixteen of twenty-eight versus eight of twenty-two, 57.1% versus 36.4%. With borrowing, the odds ratio was 1.07, with a credible interval of 0.81–1.39, but noninferiority probability 0.989 remained below the required threshold. Without borrowing, the odds ratio became 2.56, with an interval of 1.14–5.68 and noninferiority probability 0.144.^{1,2}

The COPD result was also sensitive to borrowing: the separate-group model gave an odds ratio of 1.48, with an interval of 0.67–3.09 and noninferiority probability 0.549. Noninferiority remained supported in the two larger hypoxaemic groups. In cardiogenic pulmonary oedema, the separate estimate was 0.52, with an interval of 0.29–0.91, but its superiority probability of 0.970 did not meet the specified threshold. Secondary mortality and free-day outcomes did not establish differences. Comfort scores favoured high-flow therapy, while serious adverse events affected 9.4% versus 9.9%.^{1,2}

Critique

RENOVATE addresses a common clinical decision with unusually broad recruitment. Central allocation concealment, multicentre participation, complete primary follow-up in the analysis population and explicit rescue pathways are important strengths. Its adaptive structure recognised that evidence could accumulate at different rates across diagnoses. Reporting both models was valuable because their differences affect interpretation.^{1,2}

The headline of noninferiority in four of five groups nonetheless requires qualification. Noninferiority rules out a prespecified degree of worsening under a specified model; it does not establish identical effectiveness. An odds-ratio margin of 1.55 permits a potentially important increase in intubation or death. The corresponding absolute margin changes with baseline risk and is not ten percentage points in every category. A patient or clinician might accept some trade-off for greater comfort, but the chosen margin should not be mistaken for a universally acceptable clinical loss.^{1,2}

The margin's historical justification also warrants scrutiny. It drew on an estimated effect of noninvasive ventilation against low-flow oxygen, then applied a common relative margin across different causes of respiratory failure. That requires confidence that an active-control benefit remains relevant to each enrolled population and its delivered treatment. RENOVATE did not include a concurrent low-flow group to demonstrate this. The issue is especially important in hypoxaemic disease, where the benefits of face-mask ventilation have been less consistent than in hypercapnic COPD.¹

Borrowing is neither an error nor an automatic source of validity. It can improve precision when related groups genuinely share treatment effects, and the trial's mixture model allowed particular departures for COVID-19 and cardiogenic pulmonary oedema. However, the COPD and immunocompromised groups remained linked to other hypoxaemic patients across the specified patterns. Their physiology, treatment responsiveness and direct sample sizes make those assumptions consequential. The apparent precision of the fifty-patient immunocompromised estimate is not information generated by fifty patients alone.²

The contrast between the raw immunocompromised event rates and the modelled odds ratio near one is therefore explainable statistically. It should not be labelled a transcription error. Equally, it is not sufficient to assert that an extreme small-group result must be anomalous. The no-borrowing analysis preserves a concerning signal with substantial uncertainty. Neither that post hoc analysis nor the early futility stop proves harm, but together they prevent a reassuring claim that high-flow therapy has been established as an interchangeable option in this population.^{1,2}

For COPD, only seventy-seven patients supplied direct evidence, and noninferiority was narrowly achieved in the borrowing model. Removing borrowing greatly widened uncertainty. In addition, almost one quarter of high-flow patients received rescue ventilation. This was permitted treatment rather than simple nonadherence, but it changes the clinical question: the trial tested starting with high flow while retaining access to ventilation, not replacing ventilation altogether. The supplementary escalation-or-death outcome occurred in sixteen of thirty-five versus eleven of forty-two patients, illustrating the importance of how treatment failure is defined.^{1,2}

Cardiogenic pulmonary oedema presents a different limitation. Its observed results favoured high flow, and noninferiority was supported without borrowing, but the trial excluded patients already receiving noninvasive ventilation. Most were recruited in emergency departments, and their baseline oxygenation was less impaired than in the larger hypoxaemic groups. These findings cannot establish that high flow should replace positive pressure in severe pulmonary oedema requiring immediate support. The favourable separate-group estimate also fell short of the trial's demanding superiority threshold.^{1,2}

Open treatment allocation matters because intubation is partly a clinical decision. Standard criteria reduce variation but cannot remove differences in thresholds, tolerance or co-interventions. The composite appropriately included death without intubation, avoiding an interpretation in which fewer intubations due to early deaths appear beneficial. However, combining these outcomes does not make them clinically equivalent, and the seven-day window does not capture every subsequent failure. Longer-term mortality therefore provides essential context.¹

Excluding thirty-four patients after randomisation introduces a separate concern. These were patients who died before consent could be obtained, not a random sample with lost contact details. Their exclusion was required by the permissions available to the investigators, but it narrows the intention-to-treat claim and can affect prognosis within smaller groups. Complete outcome data among the remaining 1,766 do not erase this selection.^{1,2}

The pandemic substantially changed both recruitment and the clinical environment. Separating COVID-19 was clinically defensible, and the reported blinded decision is reassuring. Nevertheless, the immunocompromised group had already stopped under an earlier classification that included some COVID-19 patients. Its final composition is consequently not identical to the population informing that stopping decision. Time adjustment addresses part of the evolving COVID-19 prognosis, but cannot make all pandemic-era care conditions equivalent.^{1,2}

Treatment exposure and patient experience also deserve attention. High-flow therapy was usually sustained for longer, whereas ventilation was intermittent despite encouragement towards continuous first-day use. RENOVAE compares these deliverable strategies, not matched-duration physiological treatments. Comfort was actually measured in the supplement and favoured high flow, although the unblinded, short-term assessments do not establish a longer-term quality-of-life benefit. The two contemporaneous editorials appropriately emphasised diagnostic heterogeneity and caution about wholesale practice change; their practical interpretations remain opinions rather than additional randomised evidence.^{1,2,3,4}

Body of Evidence

FLORALI previously compared high-flow oxygen, standard oxygen and noninvasive ventilation in 310 patients with non-hypercapnic hypoxaemic respiratory failure. Intubation occurred in 38%, 47% and 50%, respectively, without a significant overall difference. That trial raised doubts about assuming face-mask ventilation was the best active comparator in every hypoxaemic patient. RENOVAE adds a larger direct comparison, but its noninferiority design answers a different question from FLORALI's superiority analysis.⁵

In immunocompromised patients, FLORALI-IM compared high-flow oxygen alone with high flow alternating with noninvasive ventilation. Among 299 analysed patients, twenty-eight-day mortality was 36% versus 35%, an absolute difference of 1.2 percentage points, with a 95% confidence interval from -9.6 to 11.9, $P=0.83$. This did not demonstrate a mortality difference, but precision was limited. The larger direct population and different ventilation strategy should be considered alongside RENOVAE's very small, early-stopped group, rather than treating either result as definitive equivalence.⁶

A separate 2024 trial by Tan and colleagues provides relevant COPD evidence. Among 225 patients with exacerbations and moderate hypercapnic respiratory failure, treatment failure occurred in 25.7% with high-flow oxygen and 14.3% with noninvasive ventilation. Failure included either intubation or a switch to the other treatment. High flow did not meet the nine-percentage-point noninferiority margin, and intubation was more frequent, 14.2% versus 5.4%. The different failure definition is informative: rescue ventilation counts as an outcome there but can be part of a successful high-flow strategy in RENOVAE.⁷

For acute cardiogenic pulmonary oedema, Marjanovic and colleagues conducted a three-centre pilot trial with sixty analysed patients. The primary outcome was change in

respiratory rate after one hour, and the between-group difference did not reach conventional statistical significance. This small physiological comparison was insufficient to establish equivalence in intubation or mortality. RENOVATE contributes more clinical outcome information, while leaving severe presentations and the optimal selection of patients unresolved.⁸

RECOVERY-RS tested CPAP and high-flow oxygen against conventional oxygen in COVID-19. The thirty-day composite of intubation or death occurred in 36.3% versus 44.4% for CPAP against its control group, while high flow produced 44.3% versus 45.1% against its control. Only the CPAP comparison showed a significant difference. These are not equivalent to RENOVATE's face-mask bilevel ventilation comparison, and a positive comparison against oxygen does not establish superiority over every other support mode.⁹

Across these studies, both diagnosis and comparator matter. High-flow oxygen has credible advantages as a tolerable initial option, particularly in the larger nonimmunocompromised hypoxaemic populations. RENOVATE strengthens that position within its chosen noninferiority margin. It provides much less secure grounds for changing established treatment in hypercapnic COPD or reassuring clinicians about the immunocompromised subgroup. A single pooled description would conceal those distinctions.¹

High flow requires monitoring, reassessment and access to escalation. Further disease-specific trials should use clinically acceptable margins, recruit sufficient patients directly and distinguish rescue ventilation from intubation and death. These comparisons can clarify the trade-off between comfort and effectiveness.^{1,2}

Can high-flow nasal oxygen replace noninvasive ventilation across acute respiratory failure?

RENOVATE supported noninferiority in four groups under its primary model. Small samples, rescue ventilation and model sensitivity prevent a universal replacement claim, particularly in COPD and immunocompromised patients.

References

1. RENOVATE Investigators and the BRICNet Authors. High-Flow Nasal Oxygen vs Noninvasive Ventilation in Patients With Acute Respiratory Failure: The RENOVATE Randomized Clinical Trial. JAMA 2025;333(10):875–890. Published online 10 December 2024.

2. RENOvATE Investigators and the BRICNet Authors. Supplement 1 to: High-Flow Nasal Oxygen vs Noninvasive Ventilation in Patients With Acute Respiratory Failure. *JAMA* 2024. doi:10.1001/jama.2024.26244.
3. Freund Y, Vromant A. Reevaluating Respiratory Support in Acute Respiratory Failure—Insights From the RENOvATE Trial and Implications for Practice. *JAMA* 2025;333(10):848–849. Published online 10 December 2024.
4. Frat JP, Le Pape S, Thille AW. Is High-Flow Oxygen the Standard for All Patients With Acute Respiratory Failure? *JAMA* 2025;333(10):850–852. Published online 10 December 2024.
5. Frat JP, Thille AW, Mercat A, Girault C, Ragot S, Perbet S, et al.; FLORALI Study Group and REVA Network. High-Flow Oxygen through Nasal Cannula in Acute Hypoxemic Respiratory Failure. *N Engl J Med* 2015;372(23):2185–2196.
6. Coudroy R, Frat JP, Ehrmann S, Pène F, Decavèle M, Terzi N, et al.; FLORALI-IM study group and the REVA Research Network. High-flow nasal oxygen alone or alternating with non-invasive ventilation in critically ill immunocompromised patients with acute respiratory failure: a randomised controlled trial. *Lancet Respir Med* 2022;10(7):641–649.
7. Tan D, Wang B, Cao P, Wang Y, Sun J, Geng P, et al. High flow nasal cannula oxygen therapy versus non-invasive ventilation for acute exacerbations of chronic obstructive pulmonary disease with acute-moderate hypercapnic respiratory failure: a randomized controlled non-inferiority trial. *Crit Care* 2024;28:250.
8. Marjanovic N, Piton M, Lamarre J, Alleyrat C, Couvreur R, Guenezan J, et al. High-flow nasal cannula oxygen versus noninvasive ventilation for the management of acute cardiogenic pulmonary edema: a randomized controlled pilot study. *Eur J Emerg Med* 2024;31(4):267–275.
9. Perkins GD, Ji C, Connolly BA, Couper K, Lall R, Baillie JK, et al.; RECOVERY-RS Collaborators. 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.

Possible Reporting Inconsistencies. Table 3 appears to omit a minus sign before 24.9 in the credible interval for the immunocompromised group’s ninety-day mortality difference. Its primary-outcome superiority probabilities also differ from Figure 2 and supplementary eTable 3; for example, COVID-19 is listed as 0.568 rather than 0.136. Table 3 uses a COVID-19 control denominator of 437, whereas the flow diagram, primary figure, abstract and supplement consistently use 447. This chapter follows those concordant sources for the primary outcome. These discrepancies should not be confused with the legitimate differences between raw event rates and estimates obtained by Bayesian borrowing.

HOT-COVID

Nielsen FM, Klitgaard TL, Siegemund M, Laake JH, Thormar KM, Cole JM, et al.; HOT-COVID Trial Group. Lower vs Higher Oxygenation Target and Days Alive Without Life Support in COVID-19: The HOT-COVID Randomized Clinical Trial. JAMA 2024;331(14):1185–1194.

Introduction

Oxygen is fundamental supportive treatment in severe COVID-19, but the arterial oxygen tension clinicians should target remains uncertain. Raising inspired oxygen can improve measured oxygenation without necessarily improving tissue oxygen delivery or patient outcomes. It can also expose patients to high inspired oxygen fractions and may influence decisions to intubate or to continue respiratory support.^{1,2}

This is particularly relevant in COVID-19, where many patients reached ICU with profound hypoxaemia while receiving high-flow or other open-system oxygen rather than invasive ventilation. Evidence from heterogeneous, predominantly ventilated ICU populations does not by itself define the oxygen target for this setting. HOT-COVID tested whether aiming for PaO_2 of 60 rather than 90 mm Hg altered outcomes in such patients.¹

The primary outcome combines survival and freedom from major life support, so the result needs to be separated from a pure mortality effect. The intervention also remained pragmatic: the assigned target could affect escalation and withdrawal decisions as well as physiology. These are part of the clinical strategy, but make its mechanisms difficult to isolate.^{1,3}

The distinction between an oxygen target and the oxygen exposure required to achieve it is central. A higher arterial target may require substantially more inspired oxygen in severe pulmonary disease, while a lower target still needs monitoring for inadequate oxygenation. HOT-COVID therefore tested a monitored treatment policy over the ICU course. Its result cannot be reduced to an instruction to accept a particular pulse-oximeter reading without considering the patient, the measurement or the treatment being delivered.¹

Synopsis

HOT-COVID was an investigator-initiated, multicentre, parallel-group randomised trial. From August 2020 to March 2023, eleven ICUs in Denmark, Switzerland and Norway enrolled 726 adults. Although thirteen ICUs in five countries participated in preparations, no patients were enrolled in Iceland or Wales. The Danish Ministry of Education and Science funded the trial, supported by Aalborg University Hospital's Research Council.¹

Eligible patients had confirmed SARS-CoV-2 infection, an acute ICU admission, severe hypoxaemia and a functioning arterial catheter. Severe hypoxaemia meant at least ten litres per minute of pure oxygen in an open system, or invasive ventilation, noninvasive ventilation or continuous positive airway pressure irrespective of inspired oxygen fraction. Patients were expected to need supplemental oxygen for at least twenty-four hours and had to be randomised within twelve hours of ICU admission.^{1,3}

Important exclusions were chronic mechanical ventilation, home oxygen, previous bleomycin exposure, pregnancy, selected poisonings and a condition requiring hyperbaric oxygen. Patients with imminent withdrawal of active therapy, inability to obtain consent under local rules or prior HOT-ICU or HOT-COVID randomisation were excluded. Chronic dialysis was permitted.^{1,3}

Central computer-generated allocation used concealed variable-sized blocks, stratified by site. Patients, relatives and clinicians were unblinded, but analysis was undertaken without knowing which coded group represented each target. An independent data monitoring committee reviewed a planned interim analysis after 390 patients and recommended continuation.^{1,3}

Patients were allocated to PaO_2 targets of 60 or 90 mm Hg throughout ICU admission, including readmissions, for a maximum of ninety days. Inspired oxygen was adjusted using arterial blood gases and their relation to continuous pulse oximetry. At least four arterial measurements daily were expected. Values above target were permitted on room air and values below target when FiO_2 was 1.0.¹

The device, ventilator settings and other treatment were determined by treating clinicians. Open-system oxygen recipients were not to be intubated solely to obtain the allocated target when no other indication was present. The protocol also stated that target or FiO_2 alone should not determine extubation, but neither intubation nor weaning had a common detailed algorithm.¹

The primary outcome was days alive without mechanical ventilation, circulatory support or kidney replacement therapy within ninety days. Noninvasive ventilation and

nonintermittent CPAP counted as mechanical ventilation. Days between intermittent dialysis sessions counted as renal support days; ECMO itself did not. Death did not automatically produce a score of zero: patients retained preceding days alive without support.^{1,3}

High-flow nasal oxygen did not count as mechanical life support under this endpoint. Consequently, being free of the specified support did not necessarily mean breathing room air, being outside ICU or having recovered from respiratory disease. The separate hospital-free-day endpoint was relevant to those broader aspects of recovery, and the mortality endpoint was needed to examine survival directly.^{1,3}

Secondary outcomes were all-cause mortality, days alive and out of hospital, and a composite of specified serious adverse events: new shock, myocardial ischaemia, ischaemic stroke or intestinal ischaemia. The trial planned 780 participants to detect a seven-day mean primary-outcome difference with 80% power. The primary rank-based comparison was a site-stratified van Elteren test; a bootstrapped site-adjusted mean difference was supplementary. Secondary outcomes used multiplicity-adjusted 98.6% confidence intervals and $P < 0.014$.^{1,3}

The steering committee stopped recruitment at 726 participants because enrolment had become persistently slow. It had no access to accumulating outcome data. Primary outcome data were available in 351 of 365 lower-target and 346 of 361 higher-target patients; missing data reflected withdrawn or unobtainable consent. Baseline and serious-adverse-event data remained available for 362 and 358 patients respectively.^{1,3}

Median age was approximately sixty-six years and 68% of participants were male. At randomisation, approximately one quarter were invasively ventilated, around one eighth received noninvasive ventilation or CPAP, and almost two thirds used open oxygen systems. Median FiO_2 was 0.80 in each group and median SOFA score was three.¹

The strategies separated oxygen exposure. Across the intervention period, median PaO_2 was 71 mm Hg in the lower-target group and 91 mm Hg in the higher-target group; median arterial saturations were 92.9% and 95.5%, and median FiO_2 values were 0.54 and 0.68. Patients had a mean of seven arterial samples daily in each arm.^{1,3}

Median days alive without life support were eighty (IQR 9–89) versus seventy-two (IQR 2–88), $P = 0.009$. The supplementary site-adjusted mean difference was 5.8 days, 95% CI 0.2–11.5, $P = 0.04$. Additional adjustment for baseline factors gave 5.2 days, 95% CI 0.0–10.4, $P = 0.05$. The difference was driven chiefly by time without mechanical ventilation.¹

At ninety days, 106 of 351 lower-target patients and 120 of 346 higher-target patients had died, 30.2% versus 34.7%; the corrected risk ratio was 0.86, 98.6% CI 0.66–1.13, $P=0.18$. Median days alive out of hospital were fifty-nine versus forty-eight, $P=0.03$, which did not reach the adjusted secondary-outcome threshold. Serious adverse events occurred in 172 of 362 and 185 of 358 patients, 47.5% versus 51.7%, risk ratio 0.94, 98.6% CI 0.79–1.13.¹

Any mechanical ventilation was received by 70.7% versus 78.8%, and invasive ventilation by 55.2% versus 60.1%. Among participants not invasively ventilated at baseline, 110 of 272 lower-target and 133 of 276 higher-target patients later received invasive ventilation. Proning was used in 31.5% versus 41.1%. Myocardial, cerebral and intestinal ischaemic events were rare.¹

Critique

HOT-COVID is direct randomised evidence in a clinically relevant COVID-19 population. Concealed allocation, meaningful oxygen separation, multicentre delivery and blinded outcome analysis are strengths. Recruitment ended close to the planned sample because of declining case accrual rather than an apparent favourable effect, reducing concern about stopping bias.^{1,3}

The study nevertheless has limited precision for mortality and uncommon harms. Primary data were missing in 4% of randomised patients. Similar numbers were missing in both arms, but balance does not demonstrate that missing outcomes were exchangeable with observed outcomes. The absence of imputation follows the published threshold; it does not eliminate attrition uncertainty.³

The achieved lower strategy was not an arterial oxygen tension fixed at 60 mm Hg. Its median PaO_2 was 71 mm Hg and median arterial saturation was 92.9%. The evidence therefore supports an implemented lower-target strategy, not sustained acceptance of any particular low saturation. Nor did the comparator represent extreme hyperoxia: median arterial saturation was 95.5%.^{1,3}

Exposure figures derive from recorded high and low blood-gas values at twelve-hour intervals. They establish separation but are not continuous oxygen-dose measurements. Later measurements come from patients still in ICU, a changing population affected by recovery and death. They cannot show comparable cumulative exposure in the original randomised groups.^{1,3}

The primary outcome has patient importance, but its treatment of death matters. A patient who spent time without support before dying retained those days, while a survivor requiring prolonged support could have fewer. This is not a conventional death-penalised ventilator-free-day endpoint. Post hoc analyses assigning zero or minus one to death retained a difference favouring the lower target, but they use the same patients and do not provide independent confirmation.³

The rank-test P value and the bootstrapped mean difference are related but distinct analyses. The difference between the displayed medians is not an adjusted individual treatment gain. The estimated mean difference of 5.8 days is potentially useful but its confidence interval ranges from a small benefit to almost twelve days; with baseline adjustment its lower bound was zero.¹

Unblinded clinical decisions create the principal interpretive challenge. Intubation, ventilation and liberation directly contribute to the primary outcome and may respond to the assigned target. Reduced ventilation could represent appropriate avoidance of treatment, a benefit of lower oxygen exposure, different clinician thresholds, or a mixture. HOT-COVID estimates the overall strategy, not each mechanism.¹

Instructions against intubation solely to achieve a target reduce this concern but cannot remove it. Conversely, it would be wrong to dismiss a lower need for ventilation as merely bias: avoiding unnecessary ventilation can be an intended strategy effect. Mortality and ischaemic data provide reassurance about an obvious trade-off, but their intervals do not rule out clinically important benefit or harm.^{1,3}

The differences in subsequent ventilation and proning should be interpreted as events after randomisation. They may help explain how the strategy operated, but they are not baseline confounders that can simply be adjusted away to reveal a pure oxygen effect. The trial did not independently randomise intubation thresholds, proning and inspired oxygen. Nor do the lower ventilation proportions prove that all of the primary difference arose from preventing intubation; duration among ventilated patients and survival also contribute.¹

The serious-adverse-event composite was dominated by shock. The absence of a significant overall difference therefore gives limited information about the much rarer myocardial, cerebral and intestinal ischaemic events separately. This was not a safety non-inferiority trial, and no equivalence margin was established. A wide interval around an uncommon event should not be translated into certainty that accepting lower oxygenation is safe in every high-risk subgroup.¹

Per-protocol analyses were directionally consistent, but the primary per-protocol analysis of 511 patients had $P=0.12$. More restrictive adherence definitions selected much smaller subsets. These analyses are useful sensitivity checks, not estimates protected by the original randomisation.³

The shock subgroup showed a large interaction, but included only thirty-seven patients. Multiple small subgroup comparisons and imprecise estimates preclude a separate recommendation. Similarly, the absence of an interaction with recruitment date does not prove that effects were unchanged across variants or COVID-19 treatment eras.^{1,3}

The trial did not capture viral variant, COVID-specific treatment or the devices used for open-system oxygen. Its arterial-line requirement and experienced European ICUs limit generalisation to centres with different monitoring. It also does not justify extending the finding to non-COVID hypoxaemic respiratory failure.¹

Body of Evidence

HOT-ICU compared the same targets in 2,928 adults with acute hypoxaemic respiratory failure. Ninety-day mortality was 42.9% versus 42.4%, adjusted risk ratio 1.02, 95% CI 0.94–1.11. Neither mortality nor days alive without support differed significantly.⁴

A post hoc HOT-ICU subgroup included 110 patients with COVID-19 during the first pandemic wave. It favoured lower oxygenation for days alive without support, but did not demonstrate a mortality benefit or a significant interaction with the wider trial. HOT-COVID was the prospective disease-specific test prompted by that uncertain signal.⁵

PILOT tested peripheral saturation targets of 90%, 94% and 98% in 2,541 invasively ventilated patients at one academic centre. Median ventilator-free days were twenty, twenty-one and twenty-one, $P=0.81$, with similar twenty-eight-day hospital mortality. Its different intervention, setting and population constrain direct comparison with HOT-COVID.⁶

A one-year HOT-COVID follow-up was published in September 2024. Mortality was 117 of 348 versus 134 of 343, 33.6% versus 39.1%; adjusted risk ratio 0.85, 98.6% CI 0.66–1.09. Health-related quality-of-life analyses did not establish a difference. This is compatible with benefit but does not demonstrate improved long-term survival or quality of life.⁷

The accompanying editorial correctly frames oxygen as a dose-dependent intervention and notes that escalation to meet a higher target may itself be part of the strategy. That physiological rationale does not establish a universal target. HOT-COVID provides a reason to avoid routine pursuit of PaO₂ 90 mm Hg in comparable patients, alongside close assessment of work of breathing, circulation, ventilation and the illness trajectory.^{1,2}

The studies are not interchangeable votes for or against conservative oxygen. Their populations, target measurements and primary endpoints differ. HOT-COVID specifically informs patients with severe COVID-19 hypoxaemia receiving several forms of respiratory support, under frequent arterial monitoring. Its evidence is most direct for that implemented policy and less direct for an invasively ventilated patient with another diagnosis or a ward patient monitored mainly by pulse oximetry.¹

Even within COVID-19, the one-year report follows the same randomised cohort; it extends observation rather than providing an independent replication. Its value is to examine whether an early support-free-day advantage translated into durable patient benefit. As of the end of 2024, improved survival and quality of life remained unproven. The primary result can still matter clinically without implying those additional benefits.^{1,7}

Future comparisons would be easier to interpret with detailed recording of open-system devices, respiratory escalation decisions and cumulative oxygen exposure. A mortality or functional endpoint would help establish whether avoiding support changes longer-term recovery, while preserving support duration as an important outcome in its own right. These unresolved questions narrow the interpretation of HOT-COVID; they do not erase its randomised evidence that the lower-target policy increased days alive without the specified life support.¹

In adults with severe COVID-19 in ICU, should oxygen target a PaO₂ of 60 rather than 90 mm Hg?

The lower-target strategy increased days alive without life support. A mortality benefit was not established.

References

1. Nielsen FM, Klitgaard TL, Siegemund M, Laake JH, Thormar KM, Cole JM, et al.; HOT-COVID Trial Group. Lower vs Higher Oxygenation Target and Days Alive Without Life Support in COVID-19: The HOT-COVID Randomized Clinical Trial. JAMA 2024;331(14):1185–1194.

2. Schwartzstein RM. Oxygen Supplementation in COVID-19—How Much Is Enough? JAMA 2024;331(14):1177–1178.
3. HOT-COVID Trial Group. Supplemental online content to: Lower vs Higher Oxygenation Target and Days Alive Without Life Support in COVID-19. JAMA 2024. doi:10.1001/jama.2024.2934.
4. Schjørring OL, Klitgaard TL, Perner A, Wetterslev J, Lange T, Siegemund M, et al.; HOT-ICU Investigators. Lower or Higher Oxygenation Targets for Acute Hypoxemic Respiratory Failure. N Engl J Med 2021;384(14):1301–1311.
5. Rasmussen BS, Klitgaard TL, Perner A, Brand BA, Hildebrandt T, Siegemund M, et al. Oxygenation targets in ICU patients with COVID-19: A post hoc subgroup analysis of the HOT-ICU trial. Acta Anaesthesiol Scand 2022;66(1):76–84.
6. Semler MW, Casey JD, Lloyd BD, Hastings PG, Hays MA, Stollings JL, et al.; PILOT Investigators and the Pragmatic Critical Care Research Group. Oxygen-Saturation Targets for Critically Ill Adults Receiving Mechanical Ventilation. N Engl J Med 2022;387(19):1759–1769.
7. Crescioli E, Nielsen FM, Bunzel AM, Eriksen ASB, Siegemund M, Poulsen LM, et al. Long-term mortality and health-related quality of life with lower versus higher oxygenation targets in intensive care unit patients with COVID-19 and severe hypoxaemia. Intensive Care Med 2024;50:1603–1613.

PROLABI

Mascia L, Fanelli V, Mistretta A, Filippini M, Zanin M, Berardino M, et al.
Lung-protective mechanical ventilation in patients with severe acute brain injury: a multicenter randomized clinical trial (PROLABI). Am J Respir Crit Care Med
2024;210(9):1123–1131.

Introduction

Mechanical ventilation after severe acute brain injury has to protect two vulnerable organs. Reducing tidal volume may limit pulmonary stress, but maintaining carbon dioxide clearance can require a higher respiratory rate. Allowing PaCO₂ to rise can increase cerebral blood flow and intracranial pressure. Increasing positive end-expiratory pressure (PEEP) may recruit lung, yet its effects on venous return, systemic pressure and cerebral venous drainage depend on the individual patient's physiology.¹

The lungs are not necessarily normal simply because the primary diagnosis is neurological. Aspiration, pulmonary oedema, inflammation and prolonged ventilation can expose these patients to respiratory complications. However, a patient with compliant lungs and no acute respiratory distress syndrome (ARDS) may have a different balance of benefit and harm from one with established ARDS. Applying the same tidal volume and PEEP to both populations assumes a transferability that clinical outcome trials had not established.¹

The term lung protective can also obscure the intervention being tested. Tidal volume, respiratory rate, PEEP, driving pressure and gas exchange change together, and an improvement in one variable need not imply lower overall mechanical exposure. A clinical comparison therefore needs both the intended targets and the ventilation actually delivered. Cerebral monitoring is equally important if a proposed mechanism involves intracranial pressure.¹

PROLABI tested a lower-tidal-volume, higher-PEEP strategy in invasively ventilated patients with severe acute brain injury but no ARDS. The trial stopped well below its recruitment target when funding ended. Its concerning outcome difference deserves scrutiny of its direction, precision and physiological explanation; describing the study simply as negative would miss the principal result.¹

Synopsis

PROLABI was an investigator-initiated, open-label, parallel-group trial in eight Italian centres, recruiting between September 2014 and December 2018. Adults with traumatic brain injury, subarachnoid haemorrhage, intracerebral haemorrhage or ischaemic stroke were eligible if they neither obeyed commands nor opened their eyes and were expected to require at least seventy-two hours of invasive ventilation. Recruitment had to occur within twenty-four hours of ventilation after ICU admission. ARDS at screening, metabolic, toxic or anoxic coma, expected death within twenty-four hours and pregnancy were important exclusions.¹

Of 2,584 patients assessed, 190 were randomised: 104 to protective ventilation and eighty-six to conventional ventilation. Approximately 55% had traumatic brain injury, 26% subarachnoid haemorrhage, 15% intracerebral haemorrhage and 4% ischaemic stroke. Median Glasgow Coma Scale score was five, and mean age was in the early fifties. Intracranial pressure was monitored in about 38% at baseline, according to clinical judgement rather than a requirement applying to every participant.¹

Allocation used a concealed central web system, blocks of six and stratification by centre and neurological diagnosis. Both groups received volume assist-control ventilation until weaning. Protective ventilation targeted tidal volume 6 mL/kg predicted body weight and PEEP 8 cm H₂O; conventional ventilation targeted tidal volume above 8 mL/kg and PEEP 4 cm H₂O. These were two combined strategies, rather than independently randomised tidal-volume and PEEP interventions.¹

Both groups targeted PaCO₂ between 35 and 38 mm Hg and PaO₂ between 80 and 100 mm Hg. Respiratory rate was adjusted for carbon dioxide clearance. In the protective group, tidal volume could increase to 7 mL/kg if PaCO₂ exceeded 38 mm Hg despite a rate of thirty-five breaths per minute. Acute hyperventilation for intracranial hypertension remained available to clinicians. Thus, the intervention sought to limit hypercapnia rather than deliberately permit it.¹

The primary outcome was death, ventilator dependency or ARDS by day twenty-eight. Dependency meant connection to invasive mechanical ventilation. ARDS required Berlin criteria, with PaO₂/FiO₂ below 300 confirmed in three consecutive blood gases within twenty-four hours. Chest imaging, obtained when clinically indicated and technically feasible, was assessed by two investigators blinded to treatment and clinical course. Weaning used clinical assessment and spontaneous breathing trials lasting thirty to 120 minutes; being ventilator free required two consecutive calendar days of unassisted breathing under the published definitions.^{1,2}

Secondary outcomes included the individual composite components, ventilator-free days, ICU and hospital stay, neurological function at ICU discharge and six months, pneumonia, organ failure and inflammatory markers. The planned sample was 524, with an interim analysis after 262 patients. Funding ended before that interim point. The steering committee was unaware of clinical outcomes when it stopped recruitment at 190, and all randomised participants received treatment and contributed to the day-28 analysis. The University of Turin supported the study.¹

The recorded ventilation differed in several dimensions. On days two, four and six, mean tidal volumes were 6.6–7.1 mL/kg in the protective group versus approximately 8.5 mL/kg in controls. PEEP was approximately 8 versus 5–6 cm H₂O, and respiratory rate was twenty-one versus fourteen to sixteen breaths per minute. Driving pressure was lower with protective ventilation, approximately 8–9 versus 10–11 cm H₂O, but mechanical power was higher, approximately 16–20 versus 13–15 J/min. Plateau pressures were relatively low in both groups.¹

The gas-exchange targets were not uniformly achieved. Mean PaCO₂ in the protective arm ranged from 38.4 to 39.8 mm Hg at the three reported time points, compared with 36.4–37.5 mm Hg in controls. Mean PaO₂ ranged from 122 to 136 mm Hg across both arms, above the intended oxygenation range. Recorded mean intracranial pressure and cerebral perfusion pressure were broadly similar in the selectively monitored patients.¹

The primary composite occurred in 64/104 protective-arm patients and 39/86 controls, 61.5% versus 45.3%; relative risk 1.35, 95% confidence interval 1.03–1.79, P=0.025. Day-28 mortality was 30/104 versus 13/86, 28.9% versus 15.1%; relative risk 1.91, 95% confidence interval 1.06–3.42, P=0.024. Ventilator dependency occurred in 44/104 versus 24/86, relative risk 1.52, 95% confidence interval 1.01–2.28. ARDS occurred in 32/104 versus 19/86, relative risk 1.39, 95% confidence interval 0.85–2.27; that component did not show a statistically significant difference.¹

Median ventilator-free days were 3.5 versus eight, P=0.038. Six-month neurological data were available for eighty and sixty-nine patients. Death or a vegetative state occurred in 36/80 versus 20/69, 45% versus 29%; relative risk 1.55, 95% confidence interval 1.00–2.42, P=0.044. However, a separate analysis combining death ranking and functional outcome, with missing-data imputation, did not establish a significant group difference. Pneumonia assessments were missing in 40%, and baseline cytokine samples were available in 124 participants, rather than the entire randomised population.^{1,2}

Critique

PROLABI addresses an important evidence gap with concealed randomisation and complete ascertainment of the primary endpoint. Including several forms of severe brain injury improves clinical relevance, while stratification by diagnosis and centre protects against some avoidable imbalance. The ventilation targets were sufficiently distinct to change measured treatment, and blinded imaging assessment and neurological follow-up reduced detection bias for outcomes requiring judgement. These strengths remain relevant despite the small sample.¹

Early cessation substantially limits precision, but its reason matters. Recruitment ended because funding stopped, before the planned interim assessment, and the decision makers did not know the outcomes. This is not the familiar situation in which a favourable interim result triggers termination and exaggerates apparent benefit. Nevertheless, only 36% of the intended sample was recruited. Chance imbalance and instability of the effect estimate are appreciable concerns, especially with forty-three deaths overall.¹

An underpowered design does not make a statistically significant harm signal disappear. The primary confidence interval excludes no difference under the reported analysis, and its direction was accompanied by more deaths and ventilator dependency. Conversely, the mortality interval is wide and does not precisely quantify the size of harm. The accompanying editorial reported a fragility index of two. That illustrates sensitivity to a small number of event classifications; it neither proves the result is spurious nor supplies a replacement estimate.^{1,3}

The composite combines events with different importance and ascertainment. Death is objective, whereas ventilator dependency partly reflects decisions about sedation, airway protection and liberation. ARDS adds a clinical and radiological diagnosis. Agreement between imaging assessors was imperfect, with a kappa of 0.64 and a wide 95% confidence interval of 0.36–0.93. The component counts overlap and cannot be added to reconstruct the number of patients with the composite. Nor does a non-significant ARDS component establish that lung injury was unaffected.¹

The intervention is a particularly important limitation on causal interpretation. Lower tidal volume came with higher PEEP and respiratory rate, slightly higher carbon dioxide, lower driving pressure and higher mechanical power. It is impossible to identify which element, combination or downstream treatment might explain the clinical difference. Calling one arm protective is the trial's label, not evidence that every relevant pulmonary exposure was improved. Similarly, the control result does not demonstrate an isolated benefit from tidal volumes above 8 mL/kg.¹

The intended carbon dioxide range must also be distinguished from the reported means. Several protective-arm means exceeded its upper boundary, despite a substantially higher respiratory rate. Average PaO₂ exceeded the shared oxygen target in both arms. This does not invalidate the randomised strategy comparison, but it limits a claim that outcomes occurred under tightly maintained, equivalent gas-exchange conditions. Actual exposure and the workload required to deliver it matter as much as the written targets.¹

The physiological observations do not establish a cerebral mechanism. Only a minority had intracranial monitoring, and clinical selection for monitoring can identify a different risk group. Daily values were those considered most representative by the treating clinician, not a continuous record of intracranial pressure or secondary insults. Similar recorded means cannot exclude important episodes, yet hypothetical unobserved episodes cannot be invoked as proof of the mechanism either. Respiratory data were presented only through day six, whereas mortality curves separated later.¹

Other care could contribute to the strategy's overall effect or to chance imbalance. Sedation and neuromuscular blockade might respond to ventilation tolerance, while baseline neurological prognosis depends on factors extending beyond the diagnostic category and admission coma score. Their incomplete description prevents a convincing mediation analysis. These possibilities are explanations to investigate, not grounds for retrospectively attributing the deaths to a particular unmeasured factor or adjusting away the randomised intervention.^{1,3}

Longer-term neurological findings are less secure than complete day-28 outcomes. About 21% lacked six-month assessment. The available-case death-or-vegetative-state comparison is concerning, but selection into follow-up can change that estimate. The analysis ranking deaths before survivors and incorporating imputed functional outcomes did not show a significant difference. The two approaches answer related questions with different assumptions; reporting only the nominally positive comparison would overstate consistency. A worse neurological recovery among survivors specifically was not established.¹

The supplement also explores hypothetical outcomes for eighteen additional controls, matching the numerical allocation imbalance. Those people were not randomised or observed. Varying their assumed mortality demonstrates how sensitive a small trial can be, but it cannot repair randomisation, estimate missing outcomes or overturn the observed intention-to-treat result. Allocation imbalance in a small stratified study does not itself imply that eighteen controls were lost. Such scenarios should remain distinct from actual follow-up analyses.²

The practical inference is bounded. There was no demonstrated benefit from this fixed combination in severe acute brain injury without ARDS, and there was an important signal of worse outcomes. The study does not justify abandoning lung protection in patients with established ARDS, who were excluded, or assuming that higher tidal volume is intrinsically neuroprotective. It argues against treating a uniform ventilator prescription as sufficient without assessing carbon dioxide, respiratory mechanics, circulation and the injured brain.¹

Body of Evidence

The ARDS Network trial randomised 861 patients with acute lung injury or ARDS to lower or traditional tidal-volume strategies. Mortality was 31.0% versus 39.8%, with more ventilator-free days in the lower-volume arm. That established an outcome benefit in its pulmonary disease population. PROLABI enrolled patients without ARDS and with severe brain injury, so its different result does not negate the earlier trial; the indication and physiological constraints differ.⁴

The BRAINVENT physiological crossover study provides a closer neurological comparison. Twenty-seven patients were studied during low-tidal-volume ventilation at low or high PEEP. Six reached the intracranial-pressure safety threshold, requiring interruption of at least one intervention. Among those completing the interventions, average pressure changes were small and cerebral autoregulation and oxygenation did not change significantly. Carbon dioxide was controlled by adjusting respiratory rate. These short exposures demonstrate variable individual tolerance, rather than a survival benefit or the long-term safety of fixed settings.⁵

The 2020 European Society of Intensive Care Medicine consensus distinguished acute brain injury with and without clinically significant intracranial-pressure elevation. It suggested considering lung protective ventilation in patients without ARDS or significant pressure elevation, a weak recommendation with no direct supporting evidence. It could not recommend for or against that strategy when significant intracranial-pressure elevation was present. PROLABI therefore adds randomised clinical outcomes to an area previously governed largely by physiology and expert judgement; it should not be portrayed as contradicting a strong, universally applicable evidence-based recommendation.⁶

These sources address different levels of the clinical problem. An ARDS outcome trial supports treatment of established pulmonary injury. A crossover study can reveal immediate physiological intolerance under close monitoring. A consensus defines reasonable practice when outcome evidence is absent. PROLABI tests what happened

after allocation to two sustained treatment strategies in a selected neurological population. None can substitute for all the others, and their findings do not establish a single tidal volume or PEEP appropriate for every patient.¹

For a clinician facing this population at the end of 2024, the actionable lesson is to assess the consequences of ventilator changes rather than rely on their label. Reducing tidal volume while preserving carbon dioxide control can entail increased respiratory frequency and different patient–ventilator interaction. Raising PEEP requires attention to systemic and cerebral haemodynamics as well as oxygenation. PROLABI does not identify an optimal individual setting, but it makes uncritical implementation of its fixed intervention difficult to defend.¹

The next useful outcome trial would clarify the clinical population and comparator, document treatment over the period when outcomes diverge, and obtain more complete functional follow-up. Stratification by brain injury and cerebral physiological risk would improve interpretation, while repeated gas exchange and respiratory measurements could help distinguish an intended strategy from its achieved exposure. These are methodological implications of the unresolved mechanisms, rather than claims that any particular monitoring package has already been shown to improve survival.¹

Does a fixed lower-tidal-volume, higher-PEEP strategy improve outcomes after severe acute brain injury without ARDS?

PROLABI found more death, ventilator dependency or ARDS with this strategy. Its small, prematurely stopped sample warrants caution, but the harm signal should not be dismissed.

References

1. Mascia L, Fanelli V, Mistretta A, Filippini M, Zanin M, Berardino M, et al. Lung-protective mechanical ventilation in patients with severe acute brain injury: a multicenter randomized clinical trial (PROLABI). *Am J Respir Crit Care Med* 2024;210(9):1123–1131.
2. Mascia L, Fanelli V, Mistretta A, Filippini M, Zanin M, Berardino M, et al. Online data supplement to: Lung-protective mechanical ventilation in patients with severe acute brain injury: a multicenter randomized clinical trial (PROLABI). *Am J Respir Crit Care Med* 2024. doi:10.1164/rccm.202402-0375OC.

3. Taran S, Stevens RD. Does lung protective ventilation work in acute brain injury? *Am J Respir Crit Care Med*. 2024;210:1073–1075.
4. Brower RG, Matthay MA, Morris A, Schoenfeld D, Thompson BT, Wheeler A; ARDS Network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med*. 2000;342:1301–1308.
5. Beqiri E, Smielewski P, Guérin C, Czosnyka M, Robba C, Bjertnæs L, et al. Neurological and respiratory effects of lung protective ventilation in acute brain injury patients without lung injury: brain vent, a single centre randomized interventional study. *Crit Care* 2023;27:115.
6. Robba C, Poole D, McNett M, Asehnoune K, Bösel J, Bruder N, et al. Mechanical ventilation in patients with acute brain injury: recommendations of the European Society of Intensive Care Medicine consensus. *Intensive Care Med*. 2020;46:2397–2410.

Possible Reporting Inconsistencies. Table 3 gives 39/86 control composite events as 45.7%; the fraction is 45.3%, matching the abstract and results text. Its pneumonia counts, percentages and relative risk do not reconcile with the displayed denominators; 40% of pneumonia assessments were missing. Table 2 reports protective-arm mean PaCO₂ above 38 mm Hg at each displayed time point, although the discussion states that both arms remained within 35–38 mm Hg.

FAST

Burns KEA, Wong J, Rizvi L, Lafreniere-Roula M, Thorpe K, Devlin JW, et al.; Canadian Critical Care Trials Group. Frequency of screening and spontaneous breathing trial techniques: a randomized clinical trial. JAMA 2024;332(21):1808–1821.

Introduction

Liberation from invasive ventilation requires clinicians to recognise when support is no longer necessary without removing it before a patient can sustain breathing. Readiness screening and spontaneous breathing trials, or SBTs, organise this transition. Screening identifies candidates for a trial; the trial then assesses their response to a period of reduced assistance. Completing these steps promptly can matter, but neither a favourable screening result nor a successful breathing trial answers every question about extubation, including airway protection and secretion management.¹

Two practical uncertainties concern the frequency of screening and the amount of assistance provided during the SBT. Repeating a morning assessment later in the day may identify improvement earlier. It also creates additional tests, decisions and demands on staff. Pressure support and positive end-expiratory pressure can make spontaneous breathing less demanding than a T-piece, which provides oxygen without ventilator assistance. A more demanding test may reveal limited reserve, but it may also delay extubation in patients who could otherwise manage. These are competing clinical possibilities, rather than sufficient evidence for either approach.¹

FAST randomised both components together. This matters because a screening schedule operates within a pathway that includes a particular breathing test and a clinician's response to its result. The effect of extra screening could therefore differ according to how the SBT is conducted. A factorial trial can examine this dependency because both components are allocated randomly. FAST sought to determine whether screening frequency or SBT technique shortened the time to sustained liberation, while its eventual interaction finding raised a second question: how confidently can the most promising combination be identified when the average effects of the two components appear neutral?^{1,2}

Synopsis

FAST was an open-label, multicentre, 2×2 factorial randomised trial in twenty-three North American intensive care units. Its analysed population included 108 participants recruited during a 2016 pilot phase and 689 recruited between January 2018 and February 2022. Recruitment paused during the first months of the COVID-19 pandemic. Eligibility required at least twenty-four hours of invasive ventilation, the ability to initiate or trigger breaths, FiO_2 no greater than 0.70 and PEEP no greater than 12 cm H_2O . The minimum age was eighteen in the United States and sixteen in Canada. Patients were enrolled before their first equivalent SBT, so screening could identify the earliest qualifying opportunity.^{1,3}

Exclusions included ventilation for two weeks or longer, an existing tracheostomy, prior extubation during the same admission, profound neurological impairment, escalating sedation and major weakness preventing active mobilisation. Patients already receiving equivalent low-support SBT settings were also excluded. Enrolment usually occurred during weekday daytime research availability. Consent arrangements included conventional and locally approved deferred consent. There were 837 randomisations. Thirty were subsequently excluded for ineligibility or duplicate allocation and ten withdrew, leaving 797 for analysis: 198 once-daily/pressure-supported, 204 once-daily/T-piece, 195 more-frequent/pressure-supported and 200 more-frequent/T-piece. Canadian Institutes of Health Research grants funded the trial; the funders had no role in its conduct, analysis or publication decisions.^{1,3}

Independent computer-generated allocations for screening frequency and SBT technique used a central web system, randomly varying blocks and stratification by centre. Mean age was 62.4 years and 59.2% were men. Approximately four-fifths had medical admissions, and acute respiratory failure was the most common diagnosis leading to ventilation. Chronic obstructive pulmonary disease affected roughly one-quarter of participants. Median modified organ-failure scores at randomisation were five in each group. Twenty-nine patients had COVID-19. Treating staff knew the assigned strategy.¹

Once-daily screening was scheduled between 06:00 and 08:00. More-frequent screening required morning and afternoon assessments, with further assessments permitted. Patients randomised after 10:00 needed only one assessment on the first study day. Screening required adequate oxygenation, a respiratory rate no greater than thirty-five, PEEP no greater than 10 cm H_2O , a heart rate no greater than 140 and a rapid shallow breathing index below 105 breaths/min/L during a two-minute assessment without pressure support or PEEP. The index is respiratory frequency divided by tidal volume in litres.^{1,3}

Pressure-supported SBTs used pressure support greater than zero and no greater than 8 cm H₂O, with PEEP greater than zero and no greater than 5 cm H₂O. T-piece trials supplied oxygen while patients were disconnected from ventilator assistance. For COVID-19, zero-support trials on the ventilator were permitted instead of disconnection. SBTs lasted thirty to 120 minutes. Failure criteria included respiratory distress, hypoxaemia, haemodynamic deterioration and an increased requirement for vasoactive support. Failed trials prompted a return to ventilator support. Extubation was encouraged promptly after a successful SBT but remained the clinician's decision, with guidance for airway assessment and postextubation management.^{1,3}

The primary outcome was time from randomisation to unsupported spontaneous breathing sustained for at least forty-eight hours after extubation, or disconnection in patients receiving a tracheostomy after randomisation. Follow-up continued to successful liberation, intensive care discharge, death or sixty days in those remaining ventilator dependent. Cumulative incidence accounted for death as a competing event; treatment estimates came from cause-specific Cox regression. The sample-size calculation targeted 760 patients for 80% power at a two-sided 5% significance level, assuming no interaction. The published primary analysis compared the marginal effect of each component, while subsequent contrasts characterised the interaction.¹

Neither marginal comparison demonstrated faster successful extubation. More-frequent versus once-daily screening had a hazard ratio of 0.88, with a 95% confidence interval from 0.76 to 1.03 and P=0.12. Pressure-supported versus T-piece SBTs had a hazard ratio of 1.06, with a confidence interval from 0.91 to 1.23 and P=0.45. A hazard ratio above one favoured faster successful extubation for the first-named strategy. The interaction between the two allocated components was statistically significant, with P=0.009.¹

The four groups had median times to successful extubation of 2.0, 3.1, 3.9 and 2.9 days respectively, in the order listed above. Within pressure-supported SBTs, more-frequent screening was associated with slower successful extubation than once-daily screening: hazard ratio 0.70, confidence interval 0.50–0.96, P=0.02. Within once-daily screening, pressure support versus T-piece gave a hazard ratio of 1.30, confidence interval 0.98–1.70, P=0.08. The report describes Tukey adjustment for the family of four pairwise contrasts. These results did not establish that once-daily pressure support was superior to every alternative.¹

Time to the first successful SBT was shorter with pressure support than T-piece within once-daily screening, with a hazard ratio of 1.4 and P=0.007. Reintubation occurred in 16/198, 23/204, 23/195 and 17/200 patients. Ninety-day deaths numbered forty-one, thirty-nine, forty-three and thirty-six, without a clear between-strategy difference. The

once-daily/pressure-supported group also had shorter observed ventilation and lengths of stay, but these were secondary comparisons. Six-month functional and quality-of-life measurements were available in only selected surviving participants. Two serious adverse events were reported as unrelated to the interventions; other events associated with screening or SBT conduct were uncommon.^{1,3}

Critique

FAST addresses an everyday clinical process with a design suited to studying combinations. Concealed randomisation, explicit eligibility criteria and detailed recording of treatment delivery support its internal validity. Its primary outcome required sustained unsupported breathing, which is more relevant than merely passing a test. Inclusion across twenty-three units also makes it less dependent on the habits of a single service. The factorial design is a particular strength: a trial randomising only screening frequency while allowing clinicians to choose SBT technique could not identify the same interaction as reliably.^{1,2}

The interaction complicates interpretation of the marginal effects. Those estimates average each treatment across the other allocated component. A neutral average can conceal a benefit under one background strategy and a disadvantage under another. Accordingly, the two non-significant primary comparisons should not be translated into a claim that neither component can ever influence liberation. The interaction analysis is clinically relevant evidence, not a reason to discard the trial. It changes which comparisons are most useful for understanding its findings.^{1,2}

However, an unexpected interaction also demands caution. The sample size was calculated around main effects under an assumption of no interaction, not around precise comparisons between all four combinations. The authors used Tukey adjustment for the four contrasts, so describing $P=0.02$ as wholly unadjusted would be incorrect. Adjustment for that family does not eliminate uncertainty created by an unexpected pattern, limited numbers within each cell and the wider set of outcomes examined. Independent replication remains important before treating the apparent ordering as a stable clinical rule.¹

The distinction between ranking and superiority is central. Once-daily screening with pressure-supported SBTs had the shortest estimated median, but its direct comparison with once-daily T-piece did not establish a difference in successful extubation. The significant contrast was against more-frequent screening with pressure support. Nor is the gap between two reported medians automatically an adjusted treatment effect for an individual patient. A useful conclusion must retain the actual pairwise estimate and its

uncertainty, rather than label one arm unequivocally best because its median was smallest.¹

An explanation based on workflow is plausible but unproven. More frequent testing may produce more marginal or failed results, or alter a clinician's willingness to extubate late in the day. Conversely, a successful morning test may fit routine staffing and encourage prompt action. Randomisation estimates the net effect of assignment, including these behavioural responses, but does not isolate any of them. The intervention was not simply an extra measurement: it changed the sequence of opportunities to test and act. Explaining the result requires process data capable of distinguishing these pathways.^{1,2}

The early readiness assessment also complicates a simple physiological account. Both groups first underwent a brief zero-support assessment of rapid shallow breathing before the assigned SBT. The pressure-supported group was therefore not exposed exclusively to assisted breathing throughout evaluation. Meanwhile, the subsequent SBT allowed a range of support and duration. These features were shared elements of a pragmatic pathway, rather than errors, but they limit inference about one precisely specified pressure-support dose. They also help explain why other studies starting randomisation only after readiness had already been established may yield different results.^{1,3}

Death requires careful handling in an extubation outcome. The reported cause-specific hazard describes the rate of successful extubation among patients still alive and not yet successfully liberated. It is not a simple proportion eventually extubated, and it does not itself measure survival. Accounting for death in cumulative-incidence curves is appropriate, but cannot make faster extubation interchangeable with an overall survival benefit. The uncertain mortality comparisons should therefore remain separate. Similarly, shorter observed stays cannot establish better long-term recovery when subsequent functional assessment covers only a minority of the original cohort.¹

Postrandomisation exclusions removed forty of 837 allocations. Most were documented eligibility errors, with others withdrawing consent. This is a relatively small fraction, but the analysed cohort is not literally every randomised participant. The covariate-adjusted sensitivity analysis was smaller again, including 745 patients because baseline variables were missing. Agreement with the primary pattern is reassuring; it does not restore the excluded observations or independently reproduce the result. Conclusions should be grounded in the actual analysis population and should not imply complete follow-up for every endpoint.^{1,3}

Treatment fidelity was documented in useful detail, although the published summary requires attention to denominators. Major screening violations were expressed relative to screening opportunities and major SBT violations relative to SBTs, rather than as the percentage of patients affected. Additional once-daily-group screens counted as violations, whereas extra screens were permitted in the more-frequent groups. Thus, an adherence percentage is partly defined by the assigned strategy itself. Longer ventilation also creates more opportunities for tests and violations, making total test counts unsuitable as an independently randomised treatment dose.^{1,3}

Generalisability is strongest for patients entering an early liberation pathway in units where respiratory therapists and physicians coordinate care. Existing tracheostomy, profound neurological impairment, prolonged ventilation and major weakness were excluded. The study does not establish a preferred approach to those problems. It also provides no comparison with an unstructured usual-care pathway. Once-daily screening with pressure support remains a reasonable strategy, but FAST does not prove that an afternoon reassessment is harmful when a particular patient has clearly improved. The evidence supports organised assessment with clinical judgement, while the interaction deserves confirmation.^{1,2}

Body of Evidence

Ely and colleagues randomised 300 ventilated patients to a programme coupling daily screening with a two-hour spontaneous breathing trial and clinician notification, or to daily screening without those additional interventions. Ventilation lasted a median 4.5 versus six days, and the interval from meeting screening criteria to discontinuation was one versus three days. This established the value of linking identification to testing and action in that setting. It did not compare once-daily with twice-daily screening. FAST therefore asks a later question about refinement of an organised pathway, rather than contradicting the earlier finding that systematic assessment can improve on care without the same linked response.⁴

Subirà and colleagues randomised 1,153 patients considered ready for weaning to a thirty-minute pressure-supported SBT or a two-hour T-piece trial. Successful extubation occurred in 82.3% versus 74.0%, an absolute difference of 8.2 percentage points with a 95% confidence interval from 3.4 to 13.0. Reintubation was not clearly different. The comparison changed both technique and duration, so it did not isolate pressure support alone. FAST permitted thirty to 120 minutes in both technique groups and included PEEP with pressure support. It also began the allocated screening strategy before the first SBT. These differences make the interventions related but distinct.⁵

TIP-EX studied 969 patients at high risk of extubation failure, defined by older age or chronic cardiac or respiratory disease. Pressure support of 8 cm H₂O without PEEP was compared with a T-piece. Median ventilator-free days at day twenty-eight were twenty-seven in both groups, with an estimated difference of zero and a 95% confidence interval from -0.5 to one day. More pressure-support patients underwent extubation within twenty-four hours, while reintubation remained uncertain. Earlier extubation and improved ventilator-free days were therefore not interchangeable findings. This trial also differed from FAST in patient selection, absence of PEEP during pressure-supported testing and the stage at which the comparison began.⁶

A February 2024 systematic review included forty trials and 6,716 participants across several SBT comparisons. Pressure support versus T-piece increased successful extubation in sixteen trials comprising 4,462 participants, with a risk ratio of 1.07 and a 95% confidence interval from 1.04 to 1.10; certainty was moderate. The overall result for passing an SBT was less certain and sensitive to exclusion of an outlying trial. The review did not establish that every pressure-support setting, screening schedule or patient population shared the same benefit. Its search preceded publication of FAST, whose factorial interaction adds information that a pooled comparison of technique alone cannot fully address.⁷

The accompanying FAST editorial emphasised this broader methodological contribution: treatment effects can depend on the background care with which an intervention is combined. That observation helps interpret why several technically similar trials need not answer exactly the same question. It should not be used to assume a true interaction whenever trial results differ. FAST provides randomised evidence of one unexpected interaction, with an uncertain mechanism and incomplete proof of superiority across all four combinations. For practice, the collective evidence supports timely, structured assessment and makes pressure-supported SBTs a reasonable option. It provides a weaker basis for mandating one screening frequency in every unit, or delaying an otherwise appropriate extubation to comply with a rigid timetable.^{1,2}

Should more-frequent screening or pressure-supported SBTs routinely replace other liberation strategies?

Neither marginal comparison improved successful extubation. Once-daily screening with pressure support is reasonable, but the unexpected interaction does not prove it superior to every alternative or make individual afternoon reassessment inappropriate.

References

1. Burns KEA, Wong J, Rizvi L, Lafreniere-Roula M, Thorpe K, Devlin JW, et al.; Canadian Critical Care Trials Group. Frequency of screening and spontaneous breathing trial techniques: a randomized clinical trial. *JAMA* 2024;332(21):1808–1821.
2. Webb SA, Berry LR, Serpa Neto A. Ventilator weaning strategies—managing interaction between randomized treatments. *JAMA* 2024;332(21):1796–1797.
3. Burns KEA, Wong J, Rizvi L, Lafreniere-Roula M, Thorpe K, Devlin JW, et al.; Canadian Critical Care Trials Group. Supplement 4 to: Frequency of screening and spontaneous breathing trial techniques: a randomized clinical trial. *JAMA* 2024;332(21):1808–1821.
4. Ely EW, Baker AM, Dunagan DP, Burke HL, Smith AC, Kelly PT, et al. Effect on the duration of mechanical ventilation of identifying patients capable of breathing spontaneously. *N Engl J Med* 1996;335(25):1864–1869.
5. Subirà C, Hernández G, Vázquez A, Rodríguez-García R, González-Castro A, García C, et al. Effect of pressure support vs T-piece ventilation strategies during spontaneous breathing trials on successful extubation among patients receiving mechanical ventilation: a randomized clinical trial. *JAMA* 2019;321(22):2175–2182.
6. Thille AW, Gacouin A, Coudroy R, Ehrmann S, Quenot JP, Nay MA, et al.; REVA Research Network. Spontaneous-breathing trials with pressure-support ventilation or a T-piece. *N Engl J Med* 2022;387(20):1843–1854.
7. Burns KEA, Khan J, Phoophiboon V, Trivedi V, Gomez-Builes JC, Giammarioli B, et al. Spontaneous breathing trial techniques for extubating adults and children who are critically ill: a systematic review and meta-analysis. *JAMA Netw Open* 2024;7(2):e2356794.

Possible Reporting Inconsistencies. The results text gives 8,168 patients screened, while the flow diagram gives 8,571. Adherence tables label several cumulative counts per person as counts per person per day. A Table 3 footnote reverses the definition of the rapid shallow breathing index; the intervention methods correctly specify respiratory frequency divided by tidal volume.

VENT-AVOID

Duggal A, Conrad SA, Barrett NA, Saad M, Cheema T, Pannu S, et al.; VENT-AVOID Investigators. Extracorporeal Carbon Dioxide Removal to Avoid Invasive Ventilation During Exacerbations of Chronic Obstructive Pulmonary Disease: VENT-AVOID Trial – A Randomized Clinical Trial. Am J Respir Crit Care Med 2024;209(5):529–542.

Introduction

Avoiding intubation is an important objective during an exacerbation of chronic obstructive pulmonary disease (COPD), but the means of achieving it matter. Noninvasive ventilation (NIV) supports alveolar ventilation without an artificial airway. When respiratory distress or acidosis persists, however, clinicians must decide whether to continue NIV or secure the airway and provide invasive mechanical ventilation. An extracorporeal device offers a third possibility: remove carbon dioxide directly from blood, potentially reducing the ventilatory demand placed on an obstructed respiratory system. The attraction is physiological; the necessary clinical question is whether its benefits outweigh the additional intervention and its complications.¹

Extracorporeal carbon dioxide removal (ECCO₂R) uses substantially lower blood flows than conventional extracorporeal membrane oxygenation and does not provide equivalent oxygenation support. It can lower carbon dioxide and correct respiratory acidosis, but still requires vascular access, an extracorporeal circuit and anticoagulation. Bleeding, thrombocytopenia, haemolysis and technical failure are therefore relevant even when the intention is to avoid a more invasive form of respiratory support. Freedom from an endotracheal tube is not necessarily freedom from intensive treatment.^{1,2}

Two distinct indications had emerged: preventing intubation in patients thought likely to fail NIV, and facilitating extubation in patients already receiving invasive ventilation. They have different opportunities for benefit. A patient who would recover on NIV alone may gain little from extracorporeal support, whereas a patient whose extubation is genuinely prevented by ventilatory failure could conceivably benefit. Early studies established physiological feasibility but did not securely establish that either strategy improved clinically important outcomes. VENT-AVOID tested both indications within a randomised trial, asking whether adding ECCO₂R to standard care increased time alive and free of invasive ventilation during the first five days.¹

Synopsis

VENT-AVOID was an investigator-led, open-label, multicentre trial with two separately randomised strata. It was sponsored by ALung Technologies, manufacturer of the Hemolung device, and conducted in the United States under an investigational device exemption. Between February 2018 and May 2022, 5,992 patients were screened across 41 sites; 113 were randomised at 29 sites. Forty-eight entered the NIV stratum and 65 the invasive-ventilation stratum. Allocation to additional ECCO₂R or standard care alone was approximately 1:1 within each stratum, using site-based permuted blocks. Patients or their representatives provided consent. Clinicians and patients were not blinded. An independent committee assessed the seriousness and relatedness of adverse events, and a monitoring board reviewed accumulating safety data.^{1,2}

Eligible adults had COPD, asthma–COPD overlap, or an investigator-determined qualifying diagnosis during an exacerbation. For the NIV stratum, at least one hour of NIV had been delivered and clinicians judged the patient at high risk of intubation. The qualifying features included persisting severe acidosis, worsening hypercapnia or acidosis, failure of carbon dioxide to improve accompanied by substantial dyspnoea, tachypnoea, or intolerance with continuing deterioration. The report also described eligibility based on progressive clinical decompensation after starting NIV with severe acidosis. Thus, an arterial pH below 7.25 at randomisation was not the only route into the study. Patients in the invasive stratum had been ventilated for fewer than five days and had failed a spontaneous breathing trial or were considered unsuitable for a trial or extubation. This was an early-weaning population, not a cohort with weeks of established ventilator dependence.¹

Both groups received locally determined standard treatment. ECCO₂R was additional to that care, rather than a replacement for treatment of the exacerbation. A 15.5-Fr dual-lumen catheter was placed in a femoral or right internal jugular vein. The intended circuit blood flow was 350–550 mL/min, with sweep gas titrated to carbon dioxide removal up to 10 L/min. Treatment could continue for up to 14 days. The supplementary methods describe weaning from ECCO₂R after withdrawal of NIV or extubation. The comparator continued conventional NIV or invasive ventilation, with treatment decisions made according to institutional practice. Consequently, the study evaluated an extracorporeal-support strategy alongside local respiratory management, not two otherwise identical, tightly standardised weaning pathways.^{1,2}

The primary outcome was invasive ventilator-free days during the first 120 hours after randomisation, designated VFD-5. Among survivors, time was calculated in quarter-day increments; death before day five was assigned –1 rather than zero. If reintubation occurred within 48 hours of extubation, the intervening period was not credited as

ventilator-free time. Patients could therefore achieve five VFD-5 while still receiving NIV or ECCO₂R. Important additional outcomes included intubation, physiological changes, tracheostomy, ventilator-free days through day 30, discharge, mortality and adverse events. Nonserious events were collected until ICU discharge or day 30, whichever came first, and serious events through day 60.^{1,2}

The published supplementary methods describe a target of 180 participants, providing over 80% power for a proportional treatment effect of approximately 0.55. The assumptions included intubation within five days in 55% of control patients receiving NIV and VFD-5 between zero and 3.75 in 85% of invasive-ventilation controls. The principal comparison used a one-sided van Elteren test, which combines rank information while accounting for stratum. Within-stratum results included Hodges–Lehmann location-shift estimates and exact 95% confidence intervals. The principal report acknowledges that the endpoint and target sample size had changed during the study; the original 60-day endpoint became VFD-5. Recruitment was paused in May and terminated in August 2022 because of slow enrolment, aggravated by the pandemic, after 63% of the final target had been recruited.^{1,2}

Randomisation allocated 26 versus 22 patients in the NIV stratum and 32 versus 33 in the invasive stratum to ECCO₂R and control, respectively. Five intervention-assigned patients did not commence treatment. The per-protocol analysis included 108 participants, while the analysis labelled intention-to-treat included 110; neither contained all 113 randomised participants. The reported reasons for withdrawal are not fully consistent between the principal text and supplement, although the analysis counts agree. The invasive-stratum patient with bleeding during attempted cannulation was retained in the reported intention-to-treat and safety analyses.^{1,2}

Median age was approximately 63 years overall, and about two-thirds were women. In the NIV stratum, median baseline arterial pH was 7.32 with ECCO₂R and 7.29 with control; corresponding carbon dioxide tensions were 67.2 and 81.0 mmHg in Table 2. Median age was 69 versus 62 years. In the invasive stratum, median pH was 7.40 versus 7.36 and prior invasive-ventilation duration approximately 1.3 versus 1.1 days. Median intervention treatment periods were 3.08 days in the NIV stratum and 4.59 days in the invasive stratum. At 30 minutes, median circuit carbon dioxide removal was 95 and 78 mL/min, respectively, confirming that the device delivered extracorporeal gas exchange.^{1,2}

The primary per-protocol comparison did not show benefit: the stratum-adjusted one-sided P value was 0.36. In the NIV stratum, both groups had median VFD-5 of 5.00 days, with interquartile ranges also 5.00–5.00; the estimated shift was 0.00 days (95% CI

0.00–0.00). Three of 23 intervention patients and one of 22 controls were intubated within five days. In the invasive stratum, median VFD-5 was 2.00 days (IQR 0.00–3.75) versus 0.25 days (IQR 0.00–2.00), but the estimated distributional shift was 0.00 days (95% CI 0.00–1.25), with a within-stratum two-sided P value of 0.21. The reported intention-to-treat comparison was similarly nonsignificant, with stratum-adjusted $P=0.40$.¹

In NIV patients, carbon dioxide was lower with ECCO₂R at early assessments, without a significant between-group difference in pH. In invasive-ventilation patients, early pH and minute-ventilation differences were not sustained at 24 hours. The NIV intervention group had later ICU and hospital discharge and a worse VFD-30 distribution. No clear differences in these outcomes were demonstrated in the invasive stratum. Discharge analyses censored deaths. Improvements in eating, drinking and speaking in the invasive intervention group were measured at the end of treatment, a milestone defined differently between groups.^{1,2}

In the NIV stratum, five intervention patients died in hospital compared with none receiving standard care, reported as 22% versus 0% ($P=0.02$). The day-60 Kaplan–Meier mortality estimates were 43% versus 6% ($P=0.01$), with seven versus one observed deaths. In the invasive stratum, five patients in each group died in hospital; day-60 mortality estimates were 22% versus 31% ($P=0.72$). Four intervention-group hospital deaths across the two strata were adjudicated as possibly or definitely related to ECCO₂R, three involving bleeding. Serious events considered related to ECCO₂R occurred in nine NIV and 12 invasive-stratum patients. There were no intracranial haemorrhages, but a device malfunction necessitated reintubation in one patient.^{1,2}

Critique

The key finding is not simply that an under-recruited trial failed to achieve statistical significance. In the NIV stratum, the clinical event that ECCO₂R was intended to prevent was uncommon under standard care. Only one of 22 controls required intubation within five days, compared with the 55% rate assumed for the sample-size calculation. With most controls already achieving the best possible primary outcome, there was little opportunity for the device to improve VFD-5. Yet vascular access, anticoagulation and device-related complications remained additional exposures. The discrepancy between anticipated and observed risk therefore changes both statistical sensitivity and the clinical balance of the tested strategy.^{1,2}

Selection deserves scrutiny without retrospectively rewriting eligibility. Median pH in the NIV groups suggests substantially less acidaemia than would be expected in a population

uniformly approaching intubation for refractory respiratory acidosis. However, medians do not establish that no individual was severely acidotic, and the study allowed qualifying features other than a single pH threshold. The 2024 correspondence disputed the population's suitability; the investigators correctly emphasised that combined baseline values also included patients whose acidosis had already been treated with invasive ventilation. These arguments limit extrapolation to patients with different severity; they do not turn the observed control outcomes into evidence that ECCO₂R would benefit a more severely ill group. That remains a separate, unproven proposition.^{1,3,4}

Combining the strata increased recruitment opportunities but did not make the two questions interchangeable. Preventing a new episode of invasive ventilation and accelerating the end of an existing episode have different starting risks and outcome distributions. Stratification accounts for part of that structure in the primary test; it does not establish a common clinical effect. Conversely, selecting the more favourable invasive-stratum medians after a null overall result cannot establish efficacy for weaning. The point estimate for the Hodges–Lehmann shift was zero, despite the 1.75-day arithmetic difference between the displayed medians. Those are different statistical summaries, and the latter must not be substituted for the former as the estimated treatment effect.¹

The five-day window captures an immediate physiological intervention but incompletely captures its consequences. No NIV-stratum deaths occurred during those five days, although hospital and later mortality differed. Awarding –1 for early death appropriately ranks it below continuing ventilation, but does nothing to incorporate deaths after the window closes. Similarly, five days free of invasive ventilation may include five days supported by an extracorporeal circuit. The outcome therefore measures one aspect of respiratory support, not overall recovery, freedom from organ support or net patient benefit. The identical NIV medians and zero-width shift interval reflect the concentration of scores at the ceiling, not proof that clinically important benefit or harm has been excluded.^{1,2}

Post-randomisation exclusions further qualify interpretation. Restricting the principal presentation to patients who proceeded with their allocated treatment removes part of the consequence of offering an invasive device, particularly when clinical improvement, eligibility clarification or procedural complications influence treatment initiation. The additional analysis including 110 patients is more informative, but calling it intention-to-treat does not make it complete. A supplementary assumption-based analysis incorporated one further missing NIV participant and remained nonsignificant. Agreement across these analyses is reassuring about the absence of a primary efficacy

signal, while leaving the modest sample and incomplete randomised population as limitations on precise causal estimation.^{1,2}

The mortality imbalance in NIV patients is a serious safety signal, not a precise estimate of the mortality effect in all COPD exacerbations. It derives from small numbers, an additional outcome and an open-label trial with post-randomisation exclusions. Chance baseline imbalance is possible: the intervention group was older and had more reported cancer and cardiac comorbidity. Nevertheless, a fatal haemothorax after cannulation was judged definitely related to ECCO₂R, and other bleeding-associated deaths occurred across the trial. The absence of a statistically significant difference in aggregate serious adverse events does not negate those specific harms. A therapy with no demonstrated primary benefit requires persuasive safety evidence; a small trial failing to detect some safety differences does not provide it.^{1,2}

Other apparent benefits are weakened by how and when they were measured. The intervention group's dyspnoea score improved significantly from baseline, whereas the control group's did not, but comparing two within-group P values is not a test of the difference between treatments. At 24 hours, both groups had a median score of three. Patients unable to complete the assessment were assigned the worst score, and baseline scores differed, further complicating change-based interpretation. Physiological comparisons also involved varying numbers of observations. They establish that carbon dioxide removal occurred; they are less secure as evidence of an incremental symptomatic benefit.^{1,2}

The eating, drinking and speaking comparisons require particular caution. In invasive-stratum controls, treatment ended at extubation; in intervention patients, it ended at removal of the extracorporeal cannula, which was intended to follow extubation. Thus, the end-of-treatment assessment could compare patients at different stages of recovery rather than at the same elapsed time. Unequal treatment-period definitions also undermine a simple comparison of support durations. Meanwhile, discharge analyses treated death as censoring even though death prevents discharge alive. Especially where mortality differs, such estimates should not be read as an uncomplicated reduction or increase in time to successful recovery.^{1,2}

There are important strengths: random allocation, separate clinical strata, objective ascertainment of invasive-ventilation exposure, independent event review and explicit reporting of device-related complications. The manufacturer-sponsored trial also published clinically unfavourable findings. Slow recruitment across many centres and early termination limit precision and may reflect practical difficulty identifying the intended population. Limited local experience could plausibly influence cannulation

safety, but it cannot be assumed to explain away the harms. The published account documents endpoint changes but does not establish their timing relative to access to outcomes. The appropriate judgement is that VENT-AVOID does not support routine ECCO₂R in either tested setting. Its NIV findings actively raise concern; its invasive-stratum findings leave an uncertain research question, not a treatment recommendation.^{1,2}

Body of Evidence

The preceding literature explains why the intervention remained attractive, but also why physiological success required randomised evaluation. In the 2016 ECLAIR case–control study, 25 patients with hypercapnic respiratory failure refractory to NIV received venovenous ECCO₂R and were compared with 25 historical controls treated with invasive ventilation. Intubation was avoided in 14 intervention patients, but nine experienced major bleeding. Ninety-day mortality was 28% in both groups. The extracorporeal blood flow, averaging 1.3 L/min, also differed materially from the low-flow strategy in VENT-AVOID. Historical controls already selected for intubation cannot reliably establish how many intervention patients would otherwise have required it. ECLAIR demonstrated a potentially useful clinical pathway and a substantial complication burden, not a secure treatment effect against continued contemporary NIV.⁵

A physiological study by Diehl and colleagues in 2020 examined 12 invasively ventilated patients with exacerbations of COPD. Adding ECCO₂R alongside an adjustment of respiratory rate lowered median carbon dioxide tension from 68 to 49 mmHg and increased pH from 7.25 to 7.35. Intrinsic positive end-expiratory pressure and functional residual capacity did not significantly change. Work-of-breathing measurements in only five patients were numerically favourable with extracorporeal support. This was evidence that gas exchange could improve, but not that the mechanical consequences of airflow obstruction would necessarily reverse or extubation occur sooner. A small within-patient physiological comparison cannot determine survival, reliable weaning benefit or the full treatment burden.⁶

REST provided an important, though indirect, safety and efficacy comparison in 2021. It randomised 412 adults with acute hypoxaemic respiratory failure to lower tidal volumes facilitated by ECCO₂R or conventional protective ventilation. Ninety-day mortality was 41.5% versus 39.5% (risk ratio 1.05; 95% CI 0.83–1.33), with fewer ventilator-free days and more serious adverse events in the extracorporeal group. The population and therapeutic objective differed from hypercapnic COPD, so REST cannot answer VENT-AVOID's question. Nevertheless, it challenged the broader assumption that a favourable ventilatory strategy made possible by extracorporeal support would

necessarily produce net clinical benefit. Serious complications must be considered part of that strategy, not detached from its evaluation as merely technical events.⁷

The closest preceding randomised comparison was Barrett and colleagues' 2022 study. Eighteen patients with pH below 7.30 after at least one hour of NIV were allocated to ECCO₂R plus NIV or NIV alone. Median time to NIV discontinuation was seven versus 24.5 hours (P=0.004), but median extracorporeal support lasted 96 hours. ICU and hospital stays were longer with ECCO₂R. No major bleeding was observed, but only nine patients received the device, and the study could not establish a survival advantage or reliably exclude important harms. Its central lesson is that earlier removal of a mask can coexist with longer dependence on complex support. The positive primary outcome was therefore narrower than evidence of faster overall recovery.⁸

VENT-AVOID extended randomisation to a larger, multicentre COPD population and separated prevention of intubation from assistance with weaning. It did not confirm a clinically useful improvement in the primary endpoint and exposed an important mortality and bleeding concern in patients receiving NIV. Its accompanying editorial accordingly discouraged use to avoid intubation while identifying weaning as a possible direction for further study. That distinction is reasonable as a research priority, but the lack of a statistically significant mortality excess in the smaller invasive comparison cannot establish safety, just as numerically better ventilator-free days cannot establish efficacy.^{1,9}

By the end of 2024, the evidence supported a consistent distinction between technical capability and clinical value. Extracorporeal carbon dioxide removal can modify gas exchange. Whether those changes improve outcomes depends on selecting patients with a reversible limitation that conventional treatment cannot adequately address, achieving meaningful benefit beyond another support device, and avoiding offsetting harm. VENT-AVOID narrowed rather than closed the research question, but it provided no basis for routine adoption in the populations studied.¹

Should we add ECCO₂R to avoid intubation or facilitate early extubation during a COPD exacerbation?

Not routinely. VENT-AVOID showed no primary-outcome benefit and raised a serious harm signal in patients receiving NIV; benefit for facilitating extubation remains unproven.

References

1. Duggal A, Conrad SA, Barrett NA, Saad M, Cheema T, Pannu S, et al.; VENT-AVOID Investigators. Extracorporeal Carbon Dioxide Removal to Avoid Invasive Ventilation During Exacerbations of Chronic Obstructive Pulmonary Disease: VENT-AVOID Trial – A Randomized Clinical Trial. *Am J Respir Crit Care Med* 2024;209(5):529–542.
2. VENT-AVOID Investigators. Online data supplement to the VENT-AVOID trial. *Am J Respir Crit Care Med* 2024;209(5):529–542.
3. Tiruvoipati R, Kaul S, Gupta S, Haji K. VENT-AVOID Trial: Avoiding Acute Hypercapnic Respiratory Failure! *Am J Respir Crit Care Med* 2024;209(12):1513–1514.
4. Duggal A, Conrad SA, Brochard L, Brodie D, Hill NS. Reply to Tiruvoipati et al.: VENT-AVOID Trial: Avoiding Acute Hypercapnic Respiratory Failure! *Am J Respir Crit Care Med* 2024;209(12):1515.
5. Braune S, Sieweke A, Brettner F, Staudinger T, Joannidis M, Verbrugge S, et al. The feasibility and safety of extracorporeal carbon dioxide removal to avoid intubation in patients with COPD unresponsive to noninvasive ventilation for acute hypercapnic respiratory failure (ECLAIR study): multicentre case-control study. *Intensive Care Med* 2016;42(9):1437–1444.
6. Diehl JL, Piquilloud L, Vimpere D, Aissaoui N, Guerot E, Augy JL, et al. Physiological effects of adding ECCO₂R to invasive mechanical ventilation for COPD exacerbations. *Ann Intensive Care* 2020;10:126.
7. McNamee JJ, Gillies MA, Barrett NA, Perkins GD, Tunnicliffe W, Young D, et al.; REST Investigators. Effect of Lower Tidal Volume Ventilation Facilitated by Extracorporeal Carbon Dioxide Removal vs Standard Care Ventilation on 90-Day Mortality in Patients With Acute Hypoxemic Respiratory Failure: The REST Randomized Clinical Trial. *JAMA* 2021;326(11):1013–1023.
8. Barrett NA, Hart N, Daly KJR, Marotti M, Kostakou E, Carlin C, et al. A randomised controlled trial of non-invasive ventilation compared with extracorporeal carbon dioxide removal for acute hypercapnic exacerbations of chronic obstructive pulmonary disease. *Ann Intensive Care* 2022;12:36.
9. Piquilloud L. Extracorporeal Carbon Dioxide Removal in Chronic Obstructive Pulmonary Disease: It Depends on the Objective! *Am J Respir Crit Care Med* 2024;209(5):472–473.

Possible Reporting Inconsistencies. The principal text and supplementary Table E3 disagree on withdrawal reasons by stratum. This chapter uses their concordant analysis counts.

BIRDS

Richard JCM, Beloncle FM, Béduneau G, Mortaza S, Ehrmann S, Diehl JL, et al.; REVA Network. Pressure control plus spontaneous ventilation versus volume assist-control ventilation in acute respiratory distress syndrome. A randomised clinical trial. Intensive Care Med 2024;50(10):1647–1656.

Introduction

Allowing patients with acute respiratory distress syndrome (ARDS) to breathe spontaneously during mechanical ventilation presents competing objectives. Preserving respiratory activity may reduce exposure to deep sedation and paralysis and limit diaphragm inactivity. However, respiratory effort can also increase lung-distending pressure and tidal volume, particularly when vigorous patient effort coincides with ventilator insufflation. The relevant question is not simply whether spontaneous breathing is present, but whether the resulting pattern protects the lung while providing tolerable assistance. A ventilator mode that permits respiratory activity does not necessarily achieve that balance.¹

Pressure-controlled ventilation is itself a family of strategies. Some modes synchronise assisted breaths with patient effort; others alternate between pressure levels without allowing effort to trigger additional assistance. A short physiological crossover study found that the latter approach reduced average tidal volume and transpulmonary pressure compared with synchronised modes, but increased inspiratory effort. This suggested a potentially useful trade-off requiring active adjustment, rather than an intrinsically protective setting. It also cautioned against inferring lower respiratory workload from lower delivered volumes.²

BIRDS tested a pressure-controlled strategy permitting non-synchronised, unassisted spontaneous ventilation, termed PC-SV. Although delivered using a ventilator's airway pressure release ventilation (APRV) mode, it used conventional inspiratory timing and respiratory rates rather than the prolonged high-pressure phase and short release characteristic of other APRV approaches. The intention was to compare similar tidal-volume and positive end-expiratory pressure (PEEP) targets while encouraging a controlled contribution from spontaneous breathing. Both groups also received recommendations to reduce sedation and stop neuromuscular blockade after the initial stabilisation period.^{1,3}

The trial therefore addressed an integrated ventilation and sedation strategy in early moderate-to-severe ARDS. Its importance lies in testing whether a plausible physiological approach could improve survival or recovery, not merely whether clinicians could deliver it or reduce drug doses. The distinction between those outcomes is central to interpreting its findings.¹

Synopsis

BIRDS was an investigator-led, multicentre, parallel-group randomised trial conducted in 22 French ICUs, including 16 university and six non-university units. Recruitment took place from February 2013 to October 2018, with the principal report published in September 2024. The study received French Ministry of Health funding. Treatment allocation used permuted blocks stratified by centre. Caregivers could not be blinded; the report states that analyses were performed blinded to treatment assignment.¹

Adults were eligible within 48 hours of meeting criteria for moderate or severe ARDS, with a $\text{PaO}_2/\text{FiO}_2$ ratio of 200 mmHg or less and PEEP of at least 5 cm H_2O . Important exclusions included intracranial hypertension, severe chronic respiratory disease requiring home oxygen or ventilation, extreme obesity, severe burns, Child–Pugh class C liver disease, bone marrow transplantation or chemotherapy-related neutropenia, pneumothorax, an expected ventilation requirement below 48 hours and decisions to withhold life-sustaining treatment. A suitable study ventilator had to be available.^{1,3}

During the first three hours after enrolment, both groups received volume assist-control ventilation (ACV), deep sedation and paralysis. Tidal volume was set at 6 mL/kg predicted body weight. PEEP was increased stepwise towards a plateau pressure of 28 cm H_2O , and respiratory rate was adjusted towards a pH of 7.30–7.40. Oxygen was titrated towards saturation of 88–95% and PaO_2 of 55–80 mmHg. Patients then continued ACV or switched to PC-SV according to allocation.^{1,3}

In PC-SV, the lower pressure level matched the PEEP established during stabilisation. The higher pressure level was adjusted to achieve a tidal volume of 6 mL/kg predicted body weight while keeping plateau pressure at or below 28 cm H_2O . High-pressure duration was one second; low-pressure duration initially reproduced the respiratory rate used during ACV. After 24 hours, neuromuscular blocking agents were stopped, low-pressure duration was set to two seconds, and sedation and ventilation were adjusted to encourage spontaneous ventilation contributing 10–50% of total minute ventilation. Spontaneous tidal volume was targeted at no more than 6 mL/kg predicted body weight.^{1,3}

The intervention required repeated assessment, not just selection of a mode. Clinicians reviewed the ventilator's minute-ventilation trend at least three times daily and adjusted ventilation or sedation to approach the spontaneous-ventilation target. The sedation target was a Richmond Agitation–Sedation Scale (RASS) score between –2 and –3. In the ACV group, initial protective settings were maintained through the first 24 hours, after which neuromuscular blockade was also stopped and the same RASS target applied. Patient–ventilator interactions were assessed daily to guide further adjustments.^{1,3}

Shared recommendations restricted subsequent PEEP increases to refractory hypoxaemia and specified when excessive pressure or inadequate tidal volume should prompt reduction. After day three, patients with $\text{PaO}_2/\text{FiO}_2$ above 150 mmHg could undergo a PEEP-weaning assessment and subsequently transition to pressure support. Daily spontaneous breathing trials were recommended for suitable patients; extubation remained a clinical decision. Prone positioning was recommended when $\text{PaO}_2/\text{FiO}_2$ was 100 mmHg or less and did not require paralysis. Other rescue therapies were generally reserved for refractory hypoxaemia. Neuromuscular blockade could be restarted for unacceptable patient–ventilator interaction despite optimisation of sedation.^{1,3}

The primary outcome was death in hospital from any cause before day 60. Secondary outcomes included day-28 mortality, ventilator-free, organ failure-free and vasoactive drug-free days, ventilation and ICU duration, drug exposure, sedation depth, refractory hypoxaemia, rescue therapies and pneumothorax. For the specified organ-support-free outcomes, patients dying before day 28 received zero free days. Refractory hypoxaemia required PaO_2 below 55 mmHg or saturation below 88% for more than 30 minutes despite FiO_2 of at least 0.80.^{1,3}

The sample-size calculation assumed 35% control mortality and sought a ten-percentage-point absolute reduction. A total of 700 patients would provide 80% power with a two-sided alpha of 0.05. Comparisons used conventional tests for categorical and continuous outcomes, with risk ratios and confidence intervals for binary outcomes. No imputation was used for missing observations. The report described subgroup analyses by baseline oxygenation, respiratory-system compliance and radiological severity as preplanned.¹

Among 2,371 screened patients, 700 were randomised: 348 to PC-SV and 352 to ACV. One PC-SV participant withdrew consent before the strategy began, leaving 699 in the outcome analyses, all reported to have completed 60-day follow-up. Mean age was 62 years in each group, mean SOFA scores were approximately ten, and pneumonia accounted for about two-thirds of ARDS. Mean baseline $\text{PaO}_2/\text{FiO}_2$ was 136 versus 132 mmHg; the average interval since ARDS onset was approximately 18 hours.¹

In-hospital mortality before day 60 was 120/347 (34.6%) with PC-SV and 118/352 (33.5%) with ACV: risk ratio 1.03 (95% CI 0.84–1.27; P=0.77). Day-28 mortality was 25.1% versus 22.4% (risk ratio 1.12; 95% CI 0.86–1.46). Median ventilator-free days were 13 in both groups, with a reported difference of zero days (95% CI –4 to 4). Organ failure-free and vasoactive drug-free days, ventilation duration and ICU stay also did not differ significantly.¹

Adjunctive therapy for hypoxaemia during the first week was reported in 33.1% versus 41.3% (risk ratio 0.80; 95% CI 0.66–0.98; P=0.03). Prone positioning was used in 28.8% versus 37.8% (risk ratio 0.76; 95% CI 0.61–0.94). Refractory hypoxaemia occurred in 6.3% versus 9.1%, without a significant difference. Pneumothorax occurred in 16 patients in each group. Other individual rescue therapies did not differ significantly.¹

Sedative and neuromuscular blocker exposure was lower with PC-SV. Median cumulative sedative exposure during the first week was 550 versus 730 mg in the supplement's midazolam-equivalent units (P<0.01). Sedation was generally less deep, but days alive without sedation did not differ: medians 20 versus 19 (P=0.60). Tidal volumes were slightly higher with PC-SV, reaching mean values of 7.1 versus 6.7 mL/kg predicted body weight on day seven. Only approximately 4%, 43% and 39% of PC-SV patients were within the spontaneous-ventilation target on days one, three and seven, respectively. No reported subgroup comparison demonstrated a significant mortality benefit.^{1,3}

Critique

BIRDS provides a substantially more informative clinical comparison than a physiological demonstration of spontaneous breathing. Randomisation, multicentre recruitment, a patient-important primary outcome and near-complete inclusion of the randomised population are important strengths. Comparable initial tidal-volume, pressure and sedation objectives also reduce the chance that a result simply reflects an obviously less protective control strategy. Nevertheless, the correct conclusion is failure to demonstrate superiority, not proof that PC-SV and ACV are equally effective or equally safe. This was not a non-inferiority or equivalence trial.¹

The confidence interval is more informative than the large P value. A mortality risk ratio ranging from 0.84 to 1.27 remains compatible with clinically meaningful benefit or harm. The trial was powered for a large absolute improvement, from 35% to 25%, and was not designed to exclude smaller effects. Observed control mortality closely matched the planning assumption, so an unexpectedly low event rate does not explain the null result. The favourable hypotheses were simply not supported to the precision needed to

establish a mortality advantage. Likewise, identical ventilator-free-day medians do not imply identical individual recovery trajectories.¹

The treatment contrast needs careful definition. PC-SV combined a particular pressure pattern, non-synchronisation, a spontaneous-ventilation target and a dedicated adjustment algorithm. Its comparator combined ACV with recommendations for protective settings, sedation reduction and assessment of interactions. Clinicians were not randomised separately to different sedation practices within each mode. Lower drug use therefore reflects the effect of assignment to the whole strategy; it cannot establish that the ventilator mode alone was responsible. Nor does the trial compare spontaneous breathing with its complete suppression throughout the illness. Both groups could resume respiratory activity and progress to pressure support.^{1,3}

Incomplete attainment of the physiological target is a central limitation. Fewer than half of intervention patients were within the intended spontaneous-ventilation range at the reported later assessments. That may reflect persistent sedation, illness severity, excessive respiratory drive or the practical difficulty of implementing the algorithm. It means the trial estimates the effect of offering this strategy under the achieved conditions, not the effect of reliably maintaining every patient within an ideal physiological range. However, selecting patients who achieved that range and comparing their outcomes would introduce post-randomisation selection. Better target attainment cannot simply be assumed to reveal a benefit.^{1,3}

Moreover, the proportion of minute ventilation attributed to spontaneous breaths is not a direct measurement of muscular effort or regional lung stress. A modest spontaneous contribution does not guarantee comfortable breathing, and a pressure limit measured at the airway does not fully describe the patient's contribution to distending pressure. BIRDS did not establish that its bedside target consistently achieved the proposed balance between lung and diaphragm protection. The slightly larger tidal volumes with PC-SV further caution against assuming that its mode label guarantees a more protective delivered pattern. These limitations qualify the mechanism, rather than invalidating the clinical randomised comparison.^{1,3}

Sedation findings are plausible but narrower than a demonstrated improvement in recovery. RASS values showed that many patients remained more deeply sedated than the intended -2 to -3 range during the early days. Reducing cumulative exposure within that context is different from proving an advantage against consistently achieved light sedation. Converting several sedatives into a single equivalent dose also compresses drugs with different properties into one numerical measure. Most importantly, reduced exposure did not translate into more ventilator-free days, more days alive without

sedation or a shorter ICU stay. Patient comfort, cognition and longer-term functional benefit cannot be inferred from the dose comparison.^{1,3}

The lower use of rescue therapy is particularly vulnerable to clinician behaviour because allocation was visible and those decisions were discretionary. It was largely accompanied by less prone positioning. Yet neither refractory hypoxaemia nor the sampled proportions with $\text{PaO}_2/\text{FiO}_2$ below 100 mmHg differed significantly. Those observations do not prove that physiological indications were identical throughout the week, but they weaken a simple claim that less rescue treatment demonstrates less severe respiratory failure. A clinician might also be less willing to turn a spontaneously breathing, less deeply sedated patient prone. The publication itself discusses this possibility.^{1,3}

This difference matters because prone positioning has an independent evidence base for improving survival in an appropriately selected ARDS population. Avoiding a beneficial co-intervention is not inherently a benefit of the experimental strategy. Equally, it would be speculative to claim that equal proning would have produced a positive trial: the counterfactual was not randomised. The appropriate interpretation retains the observed co-intervention pattern as part of the pragmatic strategy comparison while recognising that it complicates attribution to ventilation mode alone. BIRDS does not justify withholding prone positioning because PC-SV is being used.^{1,4}

Other secondary findings require restraint. Numerous outcomes and repeated physiological comparisons were reported without a described multiplicity-adjusted testing hierarchy. Borderline P values should therefore be treated as supportive or exploratory, particularly when patient-important outcomes remain unchanged. The three subgroup classifications do not establish an effective phenotype: separate within-subgroup comparisons are not substitutes for evidence of a treatment-by-subgroup interaction. Physiological and sedation measurements also came from changing subsets of patients. Later averages describe patients still measured under the assigned strategy, not an unchanged randomised cohort followed with complete repeated measurements.^{1,3}

The primary endpoint was in-hospital death by day 60, not unrestricted all-cause mortality through that date. This is clinically important and relatively objective, but should retain its actual definition when results are quoted. Applicability is strongest to early, predominantly pneumonia-associated ARDS treated in units capable of implementing the specific mode and adjustment process. The exclusions, mandatory initial paralysis, relatively deep achieved sedation and recruitment ending in 2018 limit direct extrapolation to every subsequent practice environment. The long interval before publication changes the context in which the findings were received; it is not, by itself,

evidence of biased conduct. Overall, BIRDS supports a sedation-sparing effect of the tested strategy without establishing better survival or recovery, or a reason for routine replacement of protective ACV.^{1,3}

Body of Evidence

The physiological work helps explain both the rationale and its limits. In Rittayamai and colleagues' 2017 randomised crossover study, 13 patients completed comparisons of non-synchronised, partially synchronised and fully synchronised pressure-controlled ventilation. Non-synchronisation reduced average tidal volume and maximal transpulmonary pressure but increased inspiratory effort, with no difference in gas exchange. The short observation periods and recovering population could not establish clinical outcomes in early ARDS. BIRDS translated that trade-off into a treatment strategy; it did not merely repeat a comparison of pressure waveforms.²

ACURASYS, published in 2010, provided a different rationale for controlling early respiratory activity. It randomised 340 patients with severe ARDS to 48 hours of cisatracurium or placebo. The adjusted hazard ratio for 90-day death was 0.68 (95% CI 0.48–0.98; $P=0.04$), whereas crude mortality was 31.6% versus 40.7% ($P=0.08$). The distinction between the adjusted primary analysis and the unadjusted proportions is important. These findings encouraged early paralysis in selected patients, but did not establish that respiratory activity should remain suppressed throughout ARDS or answer which mode should facilitate its return. BIRDS incorporated an initial period of paralysis before testing that transition.^{1,5}

PROSEVA subsequently demonstrated the importance of evaluating the whole treatment environment. In 466 patients with severe ARDS, prolonged prone-positioning sessions reduced 28-day mortality from 32.8% to 16.0% compared with supine care. This was a test of positioning, not spontaneous ventilation, and its effect size should not be transferred to BIRDS. Its relevance is interpretative: less proning cannot automatically be counted as a desirable outcome when considering a new ventilation strategy. Whether lower use reflects improved physiology or changed clinical decisions must be examined alongside oxygenation and patient-important outcomes.⁴

Zhou and colleagues' 2017 single-centre trial randomised 138 patients with ARDS to APRV or conventional low-tidal-volume ventilation. Median ventilator-free days were 19 versus two ($P<0.001$), while ICU mortality was 19.7% versus 34.3% ($P=0.053$). The APRV strategy used a short release phase adjusted to expiratory flow, with 10–14 releases per minute. That differs materially from BIRDS' one-second high-pressure phase and conventional timing. The striking recovery result, lack of a conventionally significant

mortality difference, single-centre delivery and differing treatment settings warrant caution. BIRDS neither reproduces that trial nor disproves every approach called APRV. Shared equipment terminology is insufficient to establish that two interventions are clinically interchangeable.⁶

ROSE changed the sedation and paralysis context after BIRDS had finished recruitment. In 1,006 patients with moderate-to-severe ARDS, a 48-hour cisatracurium infusion with deep sedation was compared with usual care without routine blockade and with lighter sedation targets. Ninety-day in-hospital mortality was 42.5% versus 42.8%, an absolute difference of -0.3 percentage points (95% CI -6.4 to 5.9; P=0.93). This did not confirm a benefit from routine early paralysis in that setting. The comparison differed from ACURASYS and from BIRDS, so it cannot identify the best ventilation mode. It does, however, make the sedation strategy in the control group an essential part of any claim that a new mode is sedation-sparing.⁷

Against this background, BIRDS adds a comparatively large multicentre evaluation of a clearly described non-synchronised pressure-controlled strategy. It demonstrates that less sedative and neuromuscular blocker exposure can coexist with unchanged survival and organ-support outcomes. The remaining question is not resolved by choosing between unrestricted respiratory effort and universal suppression. It is whether a reproducible strategy can achieve a tolerable level of patient effort, preserve protective ventilation and improve outcomes when other effective care is maintained. By the end of 2024, BIRDS had not established that its approach met that clinical objective. Its strongest practical contribution is to separate a feasible, sedation-sparing option from an outcome-improving replacement for protective ACV.¹

Should PC-SV routinely replace volume assist-control ventilation in early moderate-to-severe ARDS?

No. BIRDS reduced sedative exposure but did not establish better survival or recovery; it does not justify routine replacement of protective ACV.

References

1. Richard JCM, Beloncle FM, Béduneau G, Mortaza S, Ehrmann S, Diehl JL, et al.; REVA Network. Pressure control plus spontaneous ventilation versus volume assist-control ventilation in acute respiratory distress syndrome. A randomised clinical trial. *Intensive Care Med* 2024;50(10):1647–1656.

2. Rittayamai N, Beloncle F, Goligher EC, Chen L, Mancebo J, Richard JCM, et al. Effect of inspiratory synchronization during pressure-controlled ventilation on lung distension and inspiratory effort. *Ann Intensive Care* 2017;7:100.
3. Richard JCM, Beloncle FM, Béduneau G, Mortaza S, Ehrmann S, Diehl JL, et al. Electronic supplementary material to the BIRDS trial. *Intensive Care Med* 2024;50(10):1647–1656. Methods and results, pp1–18.
4. Guérin C, Reignier J, Richard JC, Beuret P, Gacouin A, Boulain T, et al.; PROSEVA Study Group. Prone positioning in severe acute respiratory distress syndrome. *N Engl J Med* 2013;368(23):2159–2168.
5. Papazian L, Forel JM, Gacouin A, Penot-Ragon C, Perrin G, Loundou A, et al.; ACURASYS Study Investigators. Neuromuscular blockers in early acute respiratory distress syndrome. *N Engl J Med* 2010;363(12):1107–1116.
6. Zhou Y, Jin X, Lv Y, Wang P, Yang Y, Liang G, et al. Early application of airway pressure release ventilation may reduce the duration of mechanical ventilation in acute respiratory distress syndrome. *Intensive Care Med* 2017;43(11):1648–1659.
7. National Heart, Lung, and Blood Institute PETAL Clinical Trials Network. Early neuromuscular blockade in the acute respiratory distress syndrome. *N Engl J Med* 2019;380(21):1997–2008.



Gastrointestinal Trials

REVISE

Cook D, Deane A, Lauzier F, Zytaruk N, Guyatt G, Saunders L, et al.; REVISE Investigators. Stress Ulcer Prophylaxis during Invasive Mechanical Ventilation. N Engl J Med 2024;391(1):9–20.

Introduction

Stress ulcer prophylaxis is a familiar component of critical care prescribing, but familiarity does not establish its net benefit. Acid suppression may prevent upper gastrointestinal bleeding during severe illness, while changes in gastric acidity could encourage infection. A treatment given to many patients to prevent an uncommon complication requires reliable evidence about both effects. The relevant comparison also matters: showing that one acid suppressant prevents more bleeding than another does not establish whether either should be prescribed instead of no prophylaxis.¹

Modern critical care has changed the setting in which this question is asked. Enteral nutrition is common, and patients differ considerably in the duration of ventilation, haemodynamic instability and pre-existing gastrointestinal disease. Some already have a therapeutic indication for acid suppression; others receive it solely because they are critically ill. These groups cannot be treated as interchangeable. Equally, a short ICU course should not be assumed to have the same benefits or risks as indefinite treatment after hospital discharge.^{1,2}

The earlier SUP-ICU trial had found fewer clinically important bleeds with pantoprazole but no significant difference in its primary outcome of 90-day mortality. PEPTIC subsequently compared proton-pump inhibitor and histamine-2 receptor blocker strategies in ventilated patients. Together, these trials left an important question about the balance of bleeding prevention and survival, particularly across different severities of illness.^{3,4}

REVISE, Reevaluating the Inhibition of Stress Erosions, returned to a placebo-controlled comparison in adults receiving invasive ventilation. It separated a primary efficacy outcome, clinically important upper gastrointestinal bleeding, from a primary safety outcome, 90-day mortality. Its assessment also included bleeding defined with input from survivors and family members. This design made the central question concrete: does daily intravenous pantoprazole prevent bleeding that matters to patients, and what can a large blinded trial establish about the accompanying harms?¹

Synopsis

REVISE was an investigator-initiated, randomised, placebo-controlled trial conducted in 68 hospitals across eight countries. Recruitment ran from July 2019 to October 2023. Participating countries were Australia, Brazil, Canada, England, Kuwait, Pakistan, Saudi Arabia and the United States. Peer-reviewed public funding supported the study. Adults were eligible if they were invasively ventilated and expected to remain so beyond the calendar day following randomisation.¹

Patients with an established indication for acid suppression, a contraindication to it, prolonged ventilation before enrolment or more than one daily ICU dose of acid suppression were excluded. Other exclusions included recent upper gastrointestinal bleeding, specified severe oesophageal disease, dual antiplatelet treatment, or an antiplatelet drug combined with therapeutic anticoagulation. Previous uncomplicated gastro-oesophageal reflux or prehospital acid suppression did not automatically preclude participation. These criteria selected patients in whom prophylaxis could reasonably be withheld.^{1,2}

Allocation used a central website, variable permuted blocks and stratification by centre and prehospital acid-suppressant use. Patients received either pantoprazole 40 mg intravenously once daily or matching saline placebo. Local pharmacists prepared treatment without revealing allocation. Patients, families, clinical teams, research staff, adjudicators and biostatisticians remained blinded, including during analysis; no emergency unblinding occurred.¹

Treatment continued until invasive ventilation stopped, a new indication or contraindication arose, death occurred, or 90 days elapsed. It resumed if invasive ventilation restarted during the index ICU admission. Open-label treatment could therefore be appropriate when a new clinical indication developed. Other care remained with treating teams. The median study-drug exposure was five days, with an interquartile range of three to ten days.^{1,2}

The primary efficacy outcome was clinically important upper gastrointestinal bleeding in the ICU, including an ICU readmission during the index hospital admission, with follow-up capped at 90 days. It required overt bleeding accompanied by consequential haemodynamic or haematological changes, transfusion, or an invasive therapeutic intervention. Potential events underwent central blinded adjudication. The primary safety outcome was death from any cause at 90 days, including follow-up after hospital discharge.^{1,2}

Secondary outcomes included ventilator-associated pneumonia, *Clostridioides difficile* infection and patient-important upper gastrointestinal bleeding. The latter definition incorporated survivor and family perspectives and included overt bleeding associated with consequences such as transfusion of at least one unit, vasopressor initiation, investigation or intervention, or prolonged hospitalisation. It was broader than the primary bleeding definition. Other outcomes included renal replacement therapy, ventilation duration, length of stay and red-cell transfusion.^{1,2}

The target of 4,800 participants provided 85% power to detect a reduction in clinically important bleeding from 3.0% to 1.5%, using a two-sided significance level of 0.05. The primary analyses used Cox regression adjusted for prehospital acid suppression; mortality analysis also adjusted for baseline APACHE II score. Secondary outcomes and subgroup comparisons used multiplicity-adjusted P values. Their confidence intervals were not adjusted for multiplicity. Sensitivity analyses addressed centre effects, competing risk from death and treatment exposure.¹

The participant flow diagram shows 4,900 randomisations, with 2,453 allocated to pantoprazole and 2,447 to placebo. Seventy-nine patients were subsequently excluded, predominantly because consent was declined or withdrawn, leaving 4,821 in the trial analysis: 2,417 and 2,404 respectively. Mean age was 58 years and mean APACHE II score 21.7. Approximately three-quarters had medical admissions, 70% were receiving vasopressors and 6% renal replacement therapy at baseline. Prehospital acid suppression was recorded in approximately 23%.¹

Enteral nutrition was provided to approximately 92% during the trial. At least one study-drug dose was received by 96.5%. The reported 97.5% adherence incorporated permitted exemptions, whereas actual receipt of study drug on at least 80% of ventilation days occurred in 72.3%. Open-label proton-pump inhibitor administration classified as a deviation occurred in 4.5%; treatment for a new indication was not counted as a deviation. These measures describe different aspects of delivery.^{1,2}

Clinically important upper gastrointestinal bleeding occurred in 25/2,385 patients receiving pantoprazole (1.0%) and 84/2,377 receiving placebo (3.5%): hazard ratio 0.30, 95% confidence interval 0.19–0.47; $P < 0.001$. The reported absolute reduction was 2.5 percentage points, with a 95% confidence interval of 1.6–3.3. The corresponding number needed to treat is approximately 40, calculated from that absolute reduction over the trial's ICU bleeding follow-up.¹

Death by day 90 occurred in 696/2,390 patients (29.1%) and 734/2,379 (30.9%): hazard ratio 0.94, 95% CI 0.85–1.04; $P = 0.25$. The reported absolute mortality reduction was 1.7

percentage points, with a confidence interval from a 0.9-point increase to a 4.3-point reduction. Thus, the trial demonstrated a bleeding benefit but did not demonstrate a survival benefit.¹

Patient-important bleeding occurred in 1.5% versus 4.2%, hazard ratio 0.36 (95% CI 0.25–0.53; adjusted $P < 0.001$). Ventilator-associated pneumonia occurred in 23.2% versus 23.8%, hazard ratio 1.00 (95% CI 0.89–1.12; adjusted $P = 0.93$). *C. difficile* infection occurred in 28 versus 16 patients, corresponding to 1.2% versus 0.7% and a hazard ratio of 1.78 (95% CI 0.96–3.29; adjusted $P = 0.50$).¹

The supplement describes 124 clinically important bleeding episodes among the 109 patients with the primary outcome; recurrent episodes explain the larger event count. Transfusion of at least two red-cell units accompanied 99 episodes, and therapeutic endoscopy 40. These figures characterise the burden of bleeding, but multiple criteria could apply to the same episode. They should not be added together or interpreted as separate numbers of patients benefiting from pantoprazole.²

There were no significant differences in new renal replacement therapy, ventilation duration, length of stay or overall red-cell transfusion. Median ventilation duration was six days in both groups, and median ICU stay ten days. Sensitivity analyses generally supported the principal results; accounting for death as a competing event left the bleeding hazard ratio at 0.30. No multiplicity-adjusted subgroup interaction was significant. For mortality, hazard ratios were 0.85 below an APACHE II score of 25 and 1.04 at or above 25, with an adjusted interaction P value of 0.27.^{1,2}

Critique

REVISE provides a strong causal estimate of the effect of pantoprazole on important bleeding in its eligible population. Central allocation, matching placebo and extensive blinding reduce the opportunity for treatment knowledge to influence co-interventions, investigation or outcome classification. Multinational recruitment also makes the findings more persuasive than those of a narrowly selected single-centre study. The relatively brief exposure reflects the clinical period the intervention was intended to cover.¹

Bleeding ascertainment deserves particular attention. Whether overt blood becomes a clinically important event can depend on the attribution of hypotension or anaemia and on decisions to transfuse or intervene. Central adjudication blinded to allocation and centre reduces variability and differential interpretation. It cannot remove every difference in local investigation or treatment thresholds, but it supports confidence that

the substantial separation was not simply a consequence of knowing which drug a patient received.^{1,2}

The primary outcome was uncommon but consequential. An absolute reduction of approximately 25 events per 1,000 patients is clinically meaningful when prevention requires a short course of an established treatment. The number needed to treat of approximately 40 should nevertheless remain tied to this population and follow-up. It would be larger in a population with a lower untreated bleeding risk, assuming a similar relative effect, and cannot be transferred directly to patients ventilated only briefly or to continued prescribing on general wards.¹

The patient-important outcome supports the relevance of the bleeding benefit. It captures interventions and burdens that survivors and relatives considered meaningful, using a broader threshold than the primary definition. However, the two bleeding outcomes overlap substantially and should not be counted as independent replications. Nor does the reduction establish fewer transfusions overall, shorter ventilation or faster discharge: the trial did not demonstrate those downstream effects. Prevention of a serious complication remains valuable without requiring an effect on every other outcome.^{1,2}

The mortality result requires different language. The sample-size calculation targeted bleeding, and the trial did not specify a mortality non-inferiority margin in the published methods assessed here. A non-significant superiority test therefore does not establish equivalence or prove absence of harm. The confidence interval makes a large overall mortality increase less compatible with the data, while still allowing a modest increase and a potentially worthwhile reduction. Calling mortality a primary safety outcome does not change those statistical limits.¹

The infectious outcomes are similarly distinct. Pneumonia was frequent, and its estimate was close to no effect, with a reasonably informative confidence interval. *C. difficile* infection was rare: only 44 events occurred. Its estimate favoured placebo and the interval extended to a substantial relative increase with pantoprazole. The multiplicity-adjusted P value does not turn that interval into evidence of no risk. The defensible conclusion is that an increase was not established, while clinically relevant uncertainty remained.^{1,2}

Illness-severity subgroups invite overinterpretation. A statistically significant estimate in one stratum and a non-significant estimate in another do not establish that treatment effects differ. The relevant REVISE interaction test did not demonstrate such a difference. Conversely, a non-significant interaction is not proof of identical effects at

every severity. These data can support further assessment of a biologically plausible concern, but do not validate an APACHE II threshold as a stand-alone rule for prescribing or withholding prophylaxis.^{1,2}

The analysed population also needs accurate description. Although the abstract refers to 4,821 randomised patients, the flow diagram records 4,900 allocations followed by 79 exclusions. Most exclusions concerned consent rather than an adverse clinical trajectory, and numbers were relatively balanced, but the analysis is not literally an analysis of every randomised individual. Outcome denominators also differ because complete follow-up was unavailable for a small proportion. Available time-to-event information was used with censoring; small amounts of missing data should not be silently converted into complete ascertainment.¹

Treatment fidelity is strong when judged against the study's permitted stopping rules, but this differs from uninterrupted drug exposure. The 97.5% figure includes appropriate exemptions, while substantially fewer participants actually received study drug on at least 80% of ventilation days. The primary estimate therefore concerns assignment to this clinical strategy. An analysis restricted to patients with high actual exposure may be informative as a sensitivity check, but selection after randomisation cannot recover an unbiased effect of perfect adherence.^{1,2}

Generalisability is broad within prolonged invasive ventilation but bounded by eligibility. Patients with a therapeutic indication for acid suppression were excluded, as were several groups receiving intensive antithrombotic treatment. REVISE cannot determine whether withholding treatment is appropriate for them. Conversely, it does not directly establish benefit in unventilated patients or those likely to be extubated immediately. The cohort's high use of enteral nutrition shows that the observed benefit occurred alongside contemporary feeding practice; nutrition itself was not randomised, so this is not an estimate of its independent protective effect.^{1,2}

Follow-up also differed between the two primary outcomes. Bleeding surveillance concerned the ICU and specified ICU readmissions during the index hospital admission; it was not surveillance for every gastrointestinal bleed in every setting for 90 days. Mortality follow-up extended beyond discharge. The two estimates therefore answer related questions over different clinical observation periods. This matters when presenting absolute benefits and when considering whether the trial justifies continuing prophylaxis after critical care.^{1,2}

Approximately 36% of participants were co-enrolled in other studies. Such participation improves research efficiency, but can complicate attribution if another intervention

interacts with the study treatment. Similar co-enrolment proportions between groups and blinded allocation are reassuring. REVISE nevertheless estimates the effect of pantoprazole within the mixture of care and permitted concurrent research delivered at its participating centres; it does not test every possible interaction separately.^{1,2}

Finally, implementation should preserve the tested duration and purpose. A daily review of the prophylactic indication follows naturally from a strategy that stopped when invasive ventilation ceased, while permitting treatment for a new indication. Extending a short ICU course indefinitely would be an additional decision requiring its own justification. The practical message is a robust reduction in important bleeding during the eligible period, accompanied by careful interpretation of the more uncertain effects on survival and uncommon infections.¹

Body of Evidence

Earlier stress ulcer trials established that the choice of prophylactic agent can affect bleeding. In a 1998 trial involving 1,200 ventilated patients, ranitidine reduced clinically important upper gastrointestinal bleeding compared with sucralfate: 1.7% versus 3.8%, relative risk 0.44 (95% CI 0.21–0.92). This comparison supported the efficacy of acid suppression against an active alternative. It did not test a proton-pump inhibitor, and the absence of a placebo group meant that it could not directly answer whether modern ventilated patients should receive prophylaxis at all.⁵

SUP-ICU addressed the placebo-controlled question in 2018. It assigned 3,298 adults with an acute ICU admission and at least one bleeding risk factor to pantoprazole 40 mg intravenously daily or placebo. Ninety-day mortality, its primary outcome, was 31.1% versus 30.4%, relative risk 1.02 (95% CI 0.91–1.13). Clinically important bleeding occurred in 2.5% versus 4.2%. This combination anticipated REVISE's principal pattern: fewer important bleeds without a demonstrated overall survival advantage. SUP-ICU included a broader risk-defined ICU population, whereas REVISE specifically required invasive ventilation expected to continue beyond the following calendar day. The trials are complementary, rather than identical repetitions.^{1,3}

PEPTIC used a different comparison. Its cluster crossover design allocated ICU-level strategies favouring proton-pump inhibitors or histamine-2 receptor blockers in nearly 27,000 ventilated patients. In-hospital mortality by day 90 was 18.3% versus 17.5%, relative risk 1.05 (95% CI 1.00–1.10; $P=0.054$), while clinically important bleeding was 1.3% versus 1.8%, relative risk 0.73 (95% CI 0.57–0.92). Interpretation was complicated by treatment crossover, particularly use of proton-pump inhibitors during the histamine-blocker strategy. PEPTIC therefore raised a safety question while comparing

two approaches to acid suppression. It did not demonstrate the effect of proton-pump inhibitors against no prophylaxis, and its mortality estimate should not be treated as that missing comparison.⁴

A 2024 meta-analysis incorporating REVISE included 12 trials and 9,533 patients comparing proton-pump inhibitors with placebo or no prophylaxis. It found reduced clinically important bleeding, relative risk 0.51 (95% CI 0.34–0.76), with high-certainty evidence. The overall mortality estimate was 0.99 (95% CI 0.93–1.05), assessed as low certainty. A pooled within-trial severity analysis from SUP-ICU and REVISE suggested possible effect modification, judged to have intermediate credibility. Its mortality estimates were 0.89 (95% CI 0.80–0.98) in less severely ill patients and 1.08 (95% CI 0.96–1.20) in more severely ill patients. This raises a question worth retaining; it does not establish a proven mortality hazard in the latter group or a validated individual treatment rule.⁶

The accompanying REVISE editorial placed considerable weight on this severity question and proposed more selective prescribing in the most severely ill. That is an interpretation of the combined evidence, not a randomised comparison of an APACHE-directed prescribing strategy. Some patients whom clinicians might regard as particularly likely to benefit, including those receiving certain combinations of antiplatelet and anticoagulant treatment, were excluded from REVISE. Recommendations for those groups necessarily extend beyond its direct evidence. The distinction between reasonable clinical judgement and a demonstrated treatment interaction should remain explicit.^{1,7}

Against this background, REVISE strengthens the most consistent finding: short-course pantoprazole prevents clinically important bleeding in eligible invasively ventilated adults. Its placebo comparison, large sample and blinded adjudication make that conclusion more secure. It also clarifies the limits of the answer. Preventing bleeding did not produce a demonstrated survival or recovery advantage, and rare infectious harms remain less precisely estimated. The evidence supports treating a defined prophylactic indication and reviewing its duration, while avoiding claims that every ICU patient benefits equally or that all potential harms have been excluded.^{1,2}

Should eligible invasively ventilated adults receive pantoprazole for stress ulcer prophylaxis?

Pantoprazole reduces clinically important upper gastrointestinal bleeding, without a demonstrated survival benefit. Use should remain tied to the prophylactic indication, with review when invasive ventilation ends or the clinical situation changes.

References

1. Cook D, Deane A, Lauzier F, Zytaruk N, Guyatt G, Saunders L, et al.; REVISE Investigators. Stress Ulcer Prophylaxis during Invasive Mechanical Ventilation. *N Engl J Med* 2024;391(1):9–20.
2. Cook D, Deane A, Lauzier F, Zytaruk N, Guyatt G, Saunders L, et al. Supplementary Appendix to Stress Ulcer Prophylaxis during Invasive Mechanical Ventilation. *N Engl J Med* 2024;391(1):9–20.
3. Krag M, Marker S, Perner A, Wetterslev J, Wise MP, Schefold JC, et al.; SUP-ICU Trial Group. Pantoprazole in Patients at Risk for Gastrointestinal Bleeding in the ICU. *N Engl J Med* 2018;379(23):2199–2208.
4. Young PJ, Bagshaw SM, Forbes AB, Nichol AD, Wright SE, Bailey M, et al.; PEPTIC Investigators. Effect of Stress Ulcer Prophylaxis With Proton Pump Inhibitors vs Histamine-2 Receptor Blockers on In-Hospital Mortality Among ICU Patients Receiving Invasive Mechanical Ventilation: The PEPTIC Randomized Clinical Trial. *JAMA* 2020;323(7):616–626.
5. Cook D, Guyatt G, Marshall J, Leasa D, Fuller H, Hall R, et al.; Canadian Critical Care Trials Group. A Comparison of Sucralfate and Ranitidine for the Prevention of Upper Gastrointestinal Bleeding in Patients Requiring Mechanical Ventilation. *N Engl J Med* 1998;338(12):791–797.
6. Wang Y, Parpia S, Ge L, Heels-Ansdell D, Lai H, Abdar Esfahani M, et al. Proton-Pump Inhibitors to Prevent Gastrointestinal Bleeding—An Updated Meta-Analysis. *NEJM Evid* 2024;3(7):EVIDoa2400134.
7. Brown SM. Uncertain Answers—Proton-Pump Inhibition in the ICU. *N Engl J Med* 2024;391(1):78–79.

Possible Reporting Inconsistencies. The abstract describes 4,821 patients as randomised, whereas Figure 1 records 4,900 randomisations and 79 subsequent exclusions. Supplementary Table 11 also has incompatible totals: its headings show 155 patient-important bleeding episodes, but its legend describes 157 episodes in 138 patients; the main outcome table records 136 affected patients. The primary bleeding and mortality estimates above use the main report's outcome tables.



Nutrition Trials

PRECISE

Bels JLM, Thiessen S, van Gassel RJJ, Beishuizen A, De Bie Dekker A, Fraipont V, et al.; PRECISE study team. 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.

Introduction

Muscle wasting and prolonged disability are major consequences of critical illness. Providing more protein appears an intuitive way to support muscle preservation and recovery, particularly when ordinary feeding delivers less than prescribed. However, the capacity to use additional amino acids may change during acute illness. Delivering a larger dose is therefore not equivalent to demonstrating greater muscle synthesis, improved physical performance or better long-term health. An intervention intended to improve recovery needs to be assessed against those outcomes.¹

Protein trials are difficult to interpret when additional protein arrives with additional energy, different feeding routes or changes in other aspects of care. The effect of the protein dose then becomes difficult to separate from the accompanying intervention. PRECISE addressed this problem using blinded enteral feeds with similar energy density and different protein content. It compared targets of 2.0 and 1.3 g/kg per day during the period when patients required enteral nutrition in the ICU.^{1,2}

The trial also addressed a problem in outcome measurement. Physical tests and quality-of-life questionnaires are often completed only by survivors who can attend follow-up. If an intervention changes survival or the ability to participate, those tested may no longer be comparable between groups. PRECISE instead chose a primary health-utility measure that incorporated death, alongside separate physical and patient-reported outcomes. The resulting estimate needs to be understood as an assessment of overall health state, rather than simply muscle function among survivors.^{1,2}

By the time PRECISE appeared, EFFORT Protein had already challenged the assumption that higher protein delivery improves outcomes in ventilated patients at nutritional risk. PRECISE added a more tightly controlled feeding comparison and six months of recovery assessment. Its findings test whether increasing protein within an otherwise similar early

feeding strategy delivers the benefit that physiological reasoning and prescribing targets had suggested.^{1,3}

Synopsis

PRECISE was an investigator-initiated, multicentre, parallel-group randomised trial conducted in five Dutch and five Belgian hospitals. Recruitment ran from November 2020 to April 2023. Adults with an unplanned ICU admission were eligible if invasive ventilation began within 24 hours of admission and was expected to continue for at least three days. Exclusions included a contraindication to enteral feeding, moribund status or withholding treatment, BMI below 18 kg/m², kidney failure with a decision against dialysis, and severe hepatic encephalopathy.^{1,2}

Randomisation used an interactive web system with centre stratification and variable blocks. Four treatment labels represented two feeds, with two labels per group, to reduce the risk that identifying one label would reveal the entire allocation scheme. Feeds arrived in identical containers. Patients, care providers, investigators, outcome assessors, analysts and the independent monitoring board were blinded. Dutch sites could use deferred consent; Belgian sites sought proxy consent before enrolment. Patients regaining capacity were asked to confirm participation.¹

The standard feed contained approximately 6.3 g of protein per 100 mL and the high-protein feed 10 g per 100 mL. Energy densities were approximately 1.25 and 1.26 kcal/mL. Lipid content was similar; the higher protein content was accompanied by less carbohydrate. Both were commercially available enteral formulations supplied in masked containers. Treatment continued while enteral feeding was required in the ICU, for up to 90 days, and resumed after an ICU readmission within 48 hours when feeding remained necessary.^{1,2}

Enteral nutrition began within 48 hours of ICU admission at 25% of the calculated energy target and increased by 25% daily, aiming for 25 kcal/kg per day by day four. This corresponded to protein targets of approximately 1.3 and 2.0 g/kg per day. Admission bodyweight was used, but for BMI above 27 kg/m² the dosing weight was reduced to that corresponding to BMI 27. From day five, catch-up feeding could compensate for interrupted delivery during the previous day.¹

Protein delivery took priority over adjustment of the energy target. Feed volumes were not reduced for non-nutritional energy, such as propofol, or adjusted using indirect calorimetry. Additional enteral or parenteral protein supplements were prohibited. Supplemental parenteral nutrition was prohibited during the first seven days and

subsequently left to clinicians. Other measures to improve feeding tolerance, including prokinetic drugs, remained clinical decisions.¹

The primary outcome was EQ-5D-5L health utility measured at 30, 90 and 180 days after randomisation, with baseline adjustment. It combined responses across mobility, self-care, usual activities, pain or discomfort, and anxiety or depression using country-specific value sets. Death was assigned zero; some severe living health states could have negative values. Pre-illness baseline utility was obtained retrospectively from a proxy. Follow-up used visits where possible, with questionnaires also available by telephone, post or email.^{1,2}

Secondary outcomes included mortality through 180 days and repeated measures of physical function, muscle strength, psychological symptoms and other aspects of quality of life. Physical tests included six-minute walking distance, handgrip strength and the Medical Research Council sum score. Other outcomes included time to discharge alive, organ support, feeding intolerance and rehabilitation. No multiplicity adjustment was applied to the secondary outcomes.¹

The original sample-size calculation sought a health-utility difference of 0.06, with 80% power and two-sided alpha 0.05. The published methods describe an increase from 824 to 935 participants after higher-than-expected mortality increased the variability of the outcome. The primary analysis used a mixed-effects model accounting for repeated observations within patients and patients within centres. Missing outcome values were not imputed; validity depended on a missing-at-random assumption. Patients who died remained in the primary analysis.¹

Of 4,306 patients assessed, 935 were randomised: 465 to standard protein and 470 to high protein. The flow diagram distinguishes 2,255 excluded patients from 1,116 who were eligible but not enrolled. Mean age was approximately 62–63 years and 36% were female. Medical admissions comprised 73% of the standard group and 66% of the high-protein group. Baseline acute kidney injury was present in approximately 21–22%, and COVID-19 in 15%.¹

The primary outcome analysis included 430 standard-group and 419 high-group participants. Thirty-five and 51 respectively were lost before any primary outcome assessment, mainly because consent was withdrawn or survivors could not be contacted. Baseline utility was collected in 394 and 391. Median study-feeding duration was nine versus ten days. The supplement summarised protein delivery as a median of 1.19 versus 1.87 g/kg per day, with means of 0.95 versus 1.48. Mean nutritional energy intake was 19 kcal/kg per day in both groups.^{1,2}

Health utility over follow-up was lower with high protein: estimated mean difference -0.05 , 95% confidence interval -0.10 to -0.01 ; $P=0.031$. Observed mean scores at day 180 were 0.39 with standard protein and 0.36 with high protein, but the primary estimate came from the longitudinal model rather than that single comparison. Additional covariate adjustment produced a similar estimate of -0.05 , with a confidence interval of -0.10 to -0.00 and $P=0.038$.¹

Estimated 180-day mortality probabilities were 38% and 42%, respectively: hazard ratio 1.14 (95% CI 0.92 – 1.40 ; $P=0.22$). In the additional adjusted analysis, the mortality hazard ratio was 1.24 (95% CI 0.99 – 1.55 ; $P=0.063$). Gastrointestinal intolerance occurred in 26/465 versus 43/469 patients, odds ratio 1.76 (95% CI 1.06 – 2.92 ; $P=0.030$). Prokinetic use was also greater with high protein. Other recorded adverse events did not differ significantly.^{1,2}

There was no consistent improvement across physical or patient-reported secondary outcomes. Percentage of predicted six-minute walking distance favoured high protein by 5.55 percentage points (95% CI 0.04 – 11.07 ; nominal $P=0.048$), but only 154 and 133 participants contributed to that analysis. Time to discharge alive had a hazard ratio of 0.91 (95% CI 0.80 – 1.04); a post-hoc test placing more weight on early events gave $P=0.043$. Exploratory subgroup patterns suggested worse outcomes particularly in women and medical patients, without establishing a reliable prescribing rule.^{1,2}

Critique

PRECISE has important methodological strengths. Allocation was concealed through a central system, and blinding extended beyond clinicians to outcome assessment and analysis. Similar energy density, comparable feed volumes and documented separation in delivered protein make the comparison substantially cleaner than simply encouraging clinicians to give more nutrition. The intervention was practical enough to sustain for a median of approximately ten days, and the primary outcome addressed health after critical illness rather than a biochemical surrogate.^{1,2}

The feeding comparison nevertheless changed dietary composition, not an isolated molecule. Higher protein replaced carbohydrate while energy and lipid delivery remained similar. The causal estimate therefore concerns the two feeding strategies as delivered. It does not identify whether any adverse effect arose from amino-acid dose, carbohydrate displacement, gastrointestinal handling or another property of the formulas. Speculation about impaired autophagy or increased amino-acid oxidation may be biologically plausible, but these mechanisms were not established by the clinical outcome results.^{1,2}

The exposure data also require more care than the target doses alone suggest. Median protein delivery approached the intended targets, whereas means were appreciably lower. The supplement calculated intake on days when nutrition could be fully quantified, excluding days with oral intake and some partially observed days. Consequently, the quoted medians are not evidence that every randomised patient received 1.3 or 2.0 g/kg throughout the ICU stay. They demonstrate useful separation within the observed feeding period, including gradual escalation and interruptions.²

Energy intake was similar between groups, which reduces one major alternative explanation. However, it does not establish that the energy prescription was optimal for each patient. Prioritising protein meant that non-nutritional calories did not lead to lower feed volumes, and indirect calorimetry could not modify targets. Catch-up delivery could also produce individual days above the intended energy level. These features define the strategy tested and limit extrapolation to settings using a more conservative or individually measured energy prescription.^{1,2}

The primary outcome is a strength, but its meaning must remain explicit. Assigning zero after death retains a major patient outcome that a survivors-only questionnaire would omit. It also means that lower average utility may reflect more deaths, worse health among survivors, or both. The trial cannot be summarised as proving that every survivor had worse physical recovery with high protein. Nor is the measure a ranking in which death is necessarily the worst possible state: the country-specific tariffs assign negative values to some living states.^{1,2}

The estimated difference of -0.05 was statistically significant, but slightly smaller than the 0.06 difference used for sample-size planning. Its confidence interval ranged from a small decrement to a considerably larger one. This does not make the result clinically irrelevant, especially for an intervention intended to improve recovery. It does mean that the magnitude is uncertain and should be presented as an average effect on the chosen utility scale. A threshold used for planning is not a universal boundary between meaningful and meaningless health effects.¹

Missing follow-up remains the principal interpretative vulnerability. At least one primary outcome observation was available for 849 of 935 patients, with more complete loss in the high-protein group. A mixed-effects model uses partially observed trajectories efficiently, but it cannot make missingness harmless by itself. Its assumptions would be threatened if unobserved health status affected participation beyond what the available observations and covariates captured. Similar estimates after adjustment are reassuring, but they do not test every plausible departure from missing at random.^{1,2}

The distinction between participants contributing at least once and complete follow-up is also important. At day 180, the supplement records 394 standard-group and 393 high-group utility observations, including assigned values after death. Baseline utility was available for fewer people than the total contributing to the primary model. The published account does not fully reconcile those counts with its description of baseline adjustment. That limits independent reconstruction of the model and should discourage presenting the analysis as complete longitudinal measurement of all randomised patients.^{1,2}

Physical outcomes are more vulnerable to selection. Fewer than one-third of randomised participants contributed a six-minute walking measurement, and the most demanding tests are especially difficult for patients with severe disability. Restricting assessment to surviving participants who can perform a test can create an apparently favourable result even when overall outcomes are worse. The small nominal walking-distance advantage is therefore insufficient to establish functional benefit, particularly among numerous unadjusted secondary comparisons. Its direction does not invalidate the primary health-utility result; the analyses describe different selected populations and outcomes.^{1,2}

The mortality estimates support caution without proving excess deaths. Both principal and additionally adjusted confidence intervals included no difference, while allowing important harm. The larger adjusted estimate reflects a different model; it is not a second independent trial confirming mortality toxicity. Similarly, the gastrointestinal-intolerance finding and greater prokinetic use are coherent signals, but arise among several secondary outcomes. They reinforce the absence of a demonstrated advantage rather than establishing a complete mechanistic explanation for the utility difference.^{1,2}

The time-to-discharge analysis illustrates why analytical hierarchy matters. Its principal hazard-ratio interval crossed one, while a post-hoc test emphasising early events reached conventional significance. That later test can generate a hypothesis about the timing of recovery; it should not replace the original estimate as definitive proof of delayed discharge. The same discipline applies to apparent differences by sex, admission category, kidney injury or organ failure. Multiple exploratory comparisons, limited subgroup sizes and differing definitions across studies make a simple subgroup-based feeding prescription premature.¹

Applicability is strongest to early enteral feeding in ventilated adults resembling this cohort. Underweight patients, those unable to receive enteral nutrition and some patients with advanced kidney or liver disease were excluded. The trial did not test

nutritional rehabilitation after ICU discharge or a combined programme of higher protein and intensive exercise. It therefore argues against routine escalation to the higher target during the tested acute-care period, while leaving later recovery needs and specific unstudied populations unresolved.^{1,2}

Body of Evidence

The 2020 six-month follow-up of TARGET had already shown that meeting a larger proportion of estimated calorie requirements did not necessarily improve recovery. In 3,957 ventilated adults allocated to energy-dense or routine enteral feeding, delivered energy averaged 103% versus 69% of recommendations. Day-180 mortality was 29.6% versus 28.1%, relative risk 1.05 (95% CI 0.95–1.16), with no improvement in survivors' quality of life or measured functional outcomes. This was an energy comparison, so it cannot establish the effect of protein dose. It nevertheless challenged the broader assumption that closer achievement of nutritional targets automatically produces better long-term outcomes.⁴

A blinded feasibility trial by Chapple and colleagues, published online in 2020, showed that protein could be separated while maintaining similar energy delivery. It randomised 120 ventilated adults in six ICUs, with data available for 116. Mean protein delivery was 1.52 versus 0.99 g/kg ideal bodyweight per day. Blinding was maintained, and clinical estimates were too imprecise to determine benefit or harm. The study's contribution was methodological: a masked enteral formulation could create a meaningful protein contrast. PRECISE extended that approach to a substantially larger trial focused on recovery.^{1,5}

EFFORT Protein then provided a large pragmatic test in patients at high nutritional risk. It randomised 1,329 patients and analysed 1,301, comparing prescribed targets of at least 2.2 versus at most 1.2 g/kg per day. Clinicians were unblinded. Time to discharge alive by 60 days was not improved, hazard ratio 0.91 (95% CI 0.77–1.07). Sixty-day mortality was 34.6% versus 32.1%, relative risk 1.08 (95% CI 0.92–1.26). These findings weakened the case for routine high dosing. PRECISE's blinded, isocaloric design and longer recovery assessment provide complementary evidence, rather than simply reproducing the same intervention and endpoint.^{1,3}

A January 2024 meta-analysis included 23 trials and 3,303 patients comparing higher and lower protein with similar energy delivery. Overall mortality was not significantly different, relative risk 0.99 (95% CI 0.88–1.11), with low-certainty evidence. It identified concern about mortality among patients with acute kidney injury and a possible physical-function signal in two small studies combining higher protein with early

rehabilitation. These findings point to potential differences by clinical context, but the evidence was heterogeneous and preceded PRECISE. A combined nutrition-and-exercise comparison also cannot establish the independent effect of increasing protein.⁶

By the end of 2024, a Bayesian meta-analysis had incorporated more recent evidence. Published online on 27 December, although assigned to a 2025 issue, it included 22 trials and 4,164 patients. The median mortality relative risk was 1.01, with a 95% credible interval of 0.84–1.16. Under its specified models and vague priors, the probability of any mortality harm was 56.4%; the probability of a relative increase exceeding 5% was 32.4%. These estimates retained substantial uncertainty. They did not establish a mortality advantage for high protein, but neither did a probability only modestly above one-half prove harm.⁷

The PRECISE editorial emphasised both caution with high early protein and the persistent deficits experienced by survivors. Its discussion of differences from EFFORT Protein, including renal support and energy delivery, offers possible explanations for divergent subgroup patterns. Those explanations remain hypotheses because neither trial randomised patients to those contextual differences. The clinically useful synthesis is that protein targets need support from patient outcomes, and the acute phase should not be assumed to resemble later rehabilitation. PRECISE contributes a direct warning against expecting better recovery from routine escalation to 2.0 g/kg per day in its eligible population.^{1,8}

Should early enteral protein routinely be increased from 1.3 to 2.0 g/kg per day in ventilated adults?

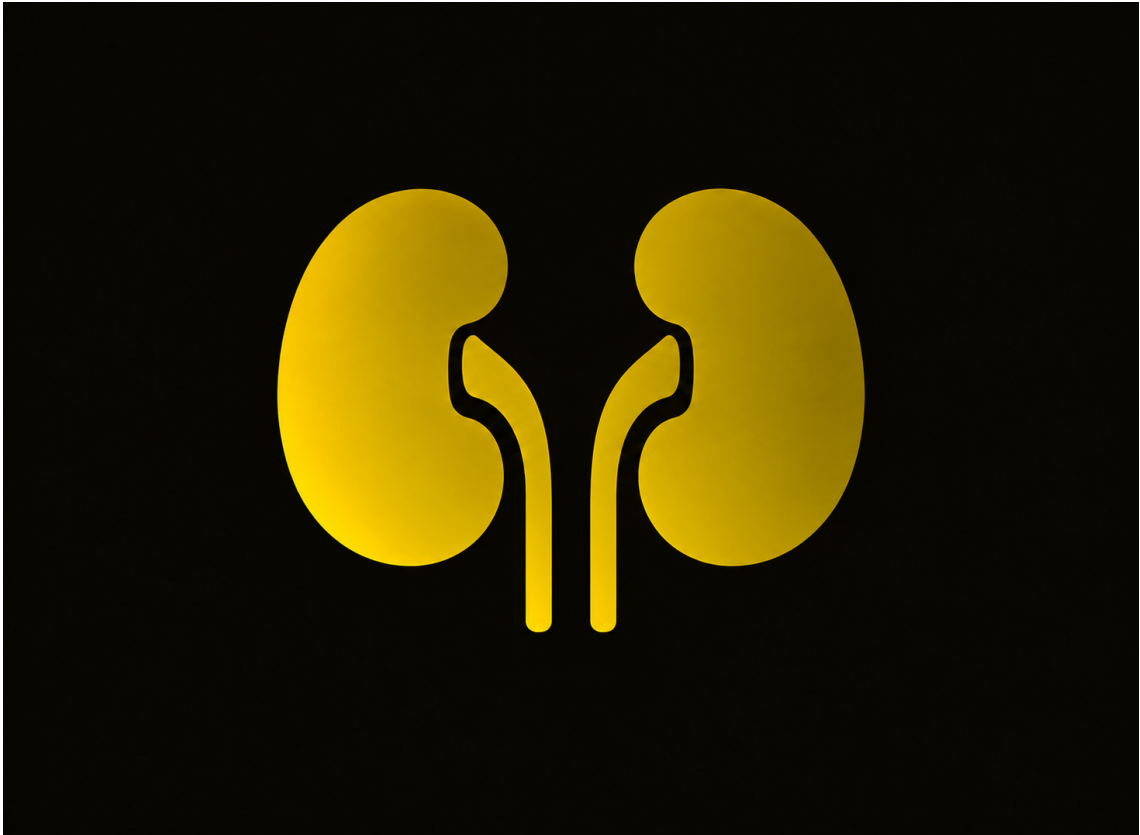
PRECISE does not support routine escalation: higher protein reduced overall health utility and did not establish a functional or survival benefit. The trial does not determine optimal protein delivery during later rehabilitation.

References

1. Bels JLM, Thiessen S, van Gassel RJJ, Beishuizen A, De Bie Dekker A, Fraipont V, et al.; PRECISE study team. 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.

2. Bels JLM, Thiessen S, van Gassel RJJ, Beishuizen A, De Bie Dekker A, Fraipont V, et al. Supplementary appendix to the PRECISE trial. *Lancet* 2024;404(10453):659–669. Published supplementary methods and results, pp1–27.
3. Heyland DK, Patel J, Compher C, Rice TW, Bear DE, Lee ZY, et al.; EFFORT Protein Trial team. 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.
4. Deane AM, Little L, Bellomo R, Chapman MJ, Davies AR, Ferrie S, et al.; TARGET Investigators. Outcomes Six Months after Delivering 100% or 70% of Enteral Calorie Requirements during Critical Illness (TARGET). A Randomized Controlled Trial. *Am J Respir Crit Care Med* 2020;201(7):814–822.
5. 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.
6. 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;28:15.
7. Heuts S, Lee ZY, Lew CCH, Bels JLM, Gabrio A, Kawczynski MJ, et al. Higher Versus Lower Protein Delivery in Critically Ill Patients: A Systematic Review and Bayesian Meta-Analysis. *Crit Care Med* 2025;53(3):e645–e655. Published online 27 December 2024.
8. Stoppe C, Hartl WH. Protein provision during critical illness. *Lancet* 2024;404(10453):630–631.

Possible Reporting Inconsistencies. The main results text describes all 3,371 non-enrolled screened patients as ineligible, whereas Figure 1 identifies 1,116 as eligible but not randomised and 2,255 as excluded. The accompanying editorial gives 435 standard-group patients; the trial's consistent allocation count is 465.



Renal Trials

PROTECTION

Landoni G, Monaco F, Ti LK, Baiardo Redaelli M, Bradic N, Comis M, et al.; PROTECTION Study Group. A Randomized Trial of Intravenous Amino Acids for Kidney Protection. N Engl J Med 2024;391(8):687–698.

Introduction

Cardiac surgery provides an unusually defined opportunity to prevent acute kidney injury. The timing of cardiopulmonary bypass is known, patients can be assessed before exposure, and an intervention can begin before the operative insult. Yet postoperative kidney dysfunction remains common, and plausible physiological treatments have often failed to produce convincing clinical benefits. A preventive treatment must therefore show more than an attractive mechanism: its effect must be demonstrated in a sufficiently large randomised comparison, with careful attention to what constitutes kidney protection.¹

Amino acids offer one such mechanism. Increased amino acid availability can recruit renal functional reserve, increasing renal blood flow and glomerular filtration. Giving an intravenous mixture around surgery might consequently attenuate the postoperative fall in filtration. However, this mechanism also complicates interpretation. If a treatment directly increases filtration, fewer patients may cross a creatinine threshold even when its effects on tubular injury, subsequent recovery or dialysis requirements are uncertain. Improvement in kidney function is potentially valuable, but the diagnostic label of acute kidney injury does not by itself establish every aspect of that value.^{1,2}

The practical question is also specific. A brief amino acid infusion during elective cardiac surgery is different from increasing nutritional protein delivery throughout a prolonged critical illness. Route, timing, dose limits, background nutrition and the patient's capacity to handle the additional substrate all matter. A successful perioperative intervention would not automatically justify high protein feeding in a general critical care population.

PROTECTION tested whether a balanced intravenous amino acid mixture, started on arrival in the operating theatre and continued for up to 72 hours, reduced postoperative acute kidney injury compared with a matched crystalloid infusion. Its scale made it an important test of functional renal protection; its outcome definitions and follow-up determine how far the result can be taken.¹

Synopsis

PROTECTION was a multicentre, randomised, double-blind, placebo-controlled trial conducted at 22 centres in Italy, Croatia and Singapore between October 2019 and January 2024. Adults undergoing elective cardiac surgery with cardiopulmonary bypass were eligible if an overnight intensive care stay was expected. Patients receiving or scheduled for kidney replacement therapy, those with an estimated glomerular filtration rate below 30 mL/min/1.73 m², kidney transplant recipients and patients with substantial liver disease or relevant amino acid contraindications were excluded. Less advanced chronic kidney disease was represented.^{1,3}

Allocation used a central web system, with computer-generated permuted blocks stratified by centre. Patients, treating clinicians, investigators, outcome assessors and the analytical team were blinded. Designated pharmacy or nursing staff preparing the infusions knew allocation but were not involved in outcome assessment. Funding came from the Italian Ministry of Health; Baxter supplied the study solutions without a reported role in the trial's conduct or interpretation. The solutions were prepared to maintain masking.^{1,3}

The intervention was a balanced amino acid mixture at a nominal dose of 2 g/kg ideal body weight per day, capped at 100 g daily. Placebo was Ringer's solution given at the same rate and volume. Infusion began on admission to the operating theatre and continued until 72 hours, intensive care discharge, commencement of kidney replacement therapy or death, whichever occurred first. The dose could be adjusted to account for amino acids provided through other nutritional routes. Other perioperative care followed local practice.^{1,3}

This was primarily a short perioperative exposure. Median infusion duration was 30 hours with amino acids and 31 hours with placebo; most patients stopped because they left intensive care. The median amino acid dose administered was 126 g in total, and median study-fluid volumes were approximately 1.26 L in both groups. Only about one fifth completed the full 72 hours. Thirty-three participants never started the assigned infusion, four crossed between groups, and some patients underwent surgery without cardiopulmonary bypass despite the intended eligibility setting.^{1,3}

The primary outcome was acute kidney injury within the first postoperative week, defined using the creatinine component of Kidney Disease: Improving Global Outcomes criteria. Stage 1 required a creatinine increase of at least 0.3 mg/dL within 48 hours or at least 1.5 times baseline within seven days. Higher stages reflected larger creatinine rises, with kidney replacement therapy included in stage 3. Urine output was omitted from the

primary definition because urinary catheters were commonly removed after 48–72 hours. A supplementary analysis incorporated available urine-output criteria.^{1,3}

Secondary outcomes included kidney injury severity, kidney replacement therapy and its duration, mechanical ventilation, intensive care and hospital stay, and mortality through 180 days. Creatinine was followed during the postoperative hospital period. The published methods allowed a preoperative baseline measurement from the preceding year, although the investigators subsequently clarified in their 2024 correspondence that all baseline samples had actually been obtained within one week before surgery. Later creatinine and quality-of-life assessments were supplementary outcomes with substantial missing data.^{1,3,4}

The planned sample of approximately 3,500 patients provided 90% power to detect a 20% relative reduction from an anticipated 25% control incidence of acute kidney injury. Three interim analyses used O'Brien–Fleming stopping boundaries. The principal comparison estimated an unadjusted relative risk; confidence intervals for additional outcomes were not adjusted for multiplicity. Thus, the primary test and the collection of secondary comparisons carried different evidential weight.¹

Of 3,512 randomised patients, one placebo recipient withdrew consent and was excluded, leaving 1,759 amino acid and 1,752 placebo participants in the reported intention-to-treat population. Median age was 66–67 years, approximately 30% were women and more than 98% were White. Baseline creatinine was about 0.95 mg/dL. Coronary and valve procedures predominated, with median bypass duration approximately 94 minutes. Two amino acid recipients lacked primary-outcome data; sensitivity analyses explored alternative handling of missing observations.^{1,3}

Acute kidney injury occurred in 474 amino acid recipients (26.9%) and 555 placebo recipients (31.7%), relative risk 0.85 (95% confidence interval 0.77–0.94; $P=0.002$). From the reported event counts, the absolute reduction was approximately 4.7 percentage points, equivalent to about 21 patients treated to prevent one creatinine-defined episode. Most episodes were stage 1. Stage 3 injury occurred in 29 patients (1.6%) versus 52 (3.0%), relative risk 0.56 (0.35–0.87), a favourable secondary finding.¹

Kidney replacement therapy was used in 24 patients (1.4%) versus 33 (1.9%), relative risk 0.73 (0.43–1.22). Its median duration was 64 hours in both groups. Median ventilation duration was 11 hours in each group, median hospital stay seven nights in each, and the analyses did not establish shorter intensive care or hospital stays. Although median intensive care stays were 30 and 34 hours, the reported comparison was a mean

difference with a confidence interval crossing zero; those medians should not be interpreted as a demonstrated four-hour benefit.¹

Mortality at 30 days was 2.8% in both groups. At 180 days, 68 patients in each group had died, relative risk 1.00 (0.72–1.38), with vital status missing in 34 and 35 patients respectively. The investigators found no substantial overall safety difference or study-drug reactions. Per-protocol and missing-data analyses supported the primary finding, and adding available urine criteria gave an acute kidney injury relative risk of 0.85 (0.77–0.94). Subgroup estimates generally pointed in the same direction, without establishing a population in which treatment was uniquely effective.^{1,3}

Critique

The trial provides persuasive evidence that this perioperative strategy reduced creatinine-defined acute kidney injury in the enrolled population. Central allocation, extensive blinding, a matched infusion and a large sample strengthen causal interpretation. The primary result was reasonably precise, with very little missing primary-outcome information, and remained similar under several sensitivity analyses. The observed absolute effect was clinically appreciable for a common postoperative complication. These strengths justify taking the functional renal result seriously rather than treating it as another small physiological pilot.^{1,3}

The crucial qualification concerns the outcome's meaning. Creatinine is a functional marker, and the intervention was designed to alter filtration. PROTECTION therefore establishes less postoperative deterioration according to a recognised functional definition, but does not establish that the same degree of structural tissue injury was prevented. No panel of tubular injury or stress biomarkers was measured. This limitation was central to the accompanying editorial and correspondence. Conversely, the absence of such measurements does not prove that the result was merely cosmetic. Both structural protection and a primarily haemodynamic benefit remain compatible with the data.^{1,2}

The stage 3 result makes a purely trivial threshold explanation less comfortable, but it does not resolve the mechanism or prove improved patient-centred outcomes. Stage 3 remained largely linked to creatinine, and it was one of several secondary comparisons. Only 81 patients experienced this outcome. The per-protocol estimate also favoured amino acids but had a confidence interval crossing one. The most defensible interpretation is a supportive signal of reduced severe functional injury, with less certainty than the primary result and a need for corroboration.^{1,3}

Dialysis and mortality results deserve the same discipline. Their point estimates do not overturn the primary finding, yet neither establishes an additional benefit. With 57 patients receiving kidney replacement therapy, the confidence interval allowed an appreciable reduction and a moderate increase. The identical observed 180-day death counts similarly do not demonstrate equivalence: the relative-risk interval extended from 0.72 to 1.38. PROTECTION was powered for a frequent creatinine-defined event, not to settle these less frequent outcomes. Describing it as a mortality-neutral or dialysis-preventing treatment would imply more certainty than the trial provides.¹

Outcome ascertainment was pragmatic but incomplete. Omitting urine output simplified diagnosis after routine catheter removal, while missing some possible oliguric episodes. The supplementary analysis was reassuring because adding the available urine information barely changed the overall effect, but it could not recreate continuous urine monitoring for the whole week. Creatinine sampling was also tied to hospital care. These features constrain the phenotype being measured rather than automatically creating treatment bias, particularly because care teams were blinded and hospital stays were similar.^{1,3}

Baseline creatinine illustrates why published correspondence matters. The broad allowable look-back window could have raised concern that old measurements misclassified postoperative change. The investigators' clarification that all baseline values were obtained within a week substantially reduces that concern. It does not supply a detailed assessment of preoperative stability for every patient, but criticism should reflect actual sampling rather than assume that year-old values were commonly used. Likewise, unmeasured albuminuria limits renal risk characterisation; its absence alone does not establish a systematic imbalance between randomised groups.^{1,3,4}

The control fluid was a substantive comparator rather than no additional intravenous volume. Matching volume helped isolate the assigned solutions' effects, and similar exposure makes simple dilution an unlikely explanation for a between-group difference. In reply to concern about chloride loading, the investigators specified a Ringer's chloride concentration of 110 mmol/L and a modest infusion rate around 40 mL/hour. It would therefore be misleading to equate this comparator with a large saline bolus. Nevertheless, the trial directly compares two infusion strategies; it does not separately estimate amino acids against usual care without the matched study infusion.^{1,4}

Dose and exposure are particularly important for translation. The 100 g/day cap means that the nominal 2 g/kg target did not translate into that weight-based amount for most larger adults. Treatment commonly lasted little more than a day, and background protein could prompt adjustment. A policy of uncapped dosing for 72 hours, or routine

continuation after intensive care discharge, would extend beyond the tested strategy. Similarly, because duration depended substantially on discharge, an observational comparison between patients receiving shorter and longer infusions would be vulnerable to differences in recovery rather than establishing a causal dose-response relationship.^{1,3}

External validity is strongest for elective cardiac surgery in centres resembling those enrolled. Severe chronic kidney disease, existing dialysis, transplantation and substantial liver disease were excluded; most participants were White, and women constituted fewer than one third. These features do not imply that the intervention fails elsewhere, but they limit direct reassurance in patients with very low renal reserve or different metabolic constraints. The broadly similar subgroup estimates cannot fill these gaps. Statistical significance within one subgroup and its absence in another would not itself establish a treatment interaction.^{1,3}

Finally, later renal recovery was incompletely measured. At 180 days, creatinine was unavailable for roughly two thirds of participants, while the quality-of-life assessment was also missing in a substantial proportion. Similar values among those assessed cannot demonstrate preserved long-term kidney health across the original cohort. An implementation decision must therefore weigh a fairly secure reduction in early functional injury against unresolved effects on dialysis, durable recovery and survival, alongside the practical requirements of an additional infusion. Those uncertainties define useful next studies without negating the result already obtained.^{1,3}

Body of Evidence

The preceding clinical evidence distinguished physiological activity from definitive renal protection. In a 2015 trial across 16 Australian and New Zealand intensive care units, Doig and colleagues randomised 474 critically ill adults to intravenous amino acid supplementation or standard care. The primary outcome, duration of renal dysfunction, did not differ. Estimated filtration and urine output increased, demonstrating biological activity. Kidney replacement therapy was numerically more frequent with amino acids, although the difference was not statistically persuasive and diminished after adjustment for baseline imbalance. This study supports the mechanism while cautioning against assuming that improved filtration necessarily produces better clinical renal outcomes in a general critical care population.⁵

A subsequent cardiac surgery pilot by Pu and colleagues randomised 69 patients to standard care or amino acids started after anaesthetic induction. The duration of renal dysfunction did not improve, but secondary measures suggested shorter acute kidney

injury, higher estimated filtration and greater urine output. Its size and multiple renal measures made it an encouraging exploratory study rather than a sufficient basis for routine adoption. It also shows why PROTECTION's contribution was more than repeating a physiological observation: the larger trial could assess a clinically recognised postoperative kidney injury outcome with substantially greater precision.⁶

Kazawa and colleagues added a small, independent 2024 trial in aortic surgery. Sixty-six patients were randomised, with one control participant withdrawing consent. Amino acids, 60 g/day for up to three days, were compared with standard care. Acute kidney injury occurred in 10 of 33 intervention patients and 18 of 32 controls; the adjusted hazard ratio was 0.44 (0.20–0.95). This supported a perioperative effect in a setting with a high control event rate. However, the single-centre sample, different regimen and limited number of events left considerable uncertainty about its magnitude and wider applicability.⁷

A 2024 meta-analysis by Pruna and colleagues identified 15 controlled perioperative studies involving 4,544 patients; six studies contributed to its acute kidney injury analysis. The pooled relative risk was 0.66 (0.47–0.94), with moderate heterogeneity. This places PROTECTION within a generally favourable perioperative literature. It should not be treated as wholly independent confirmation, because PROTECTION contributed most of the patients to the relevant evidence base. Moreover, a random-effects pooled estimate need not reproduce the simple ratio of total event proportions, and variation between small studies and the large trial argues against assuming a uniform one-third relative reduction in routine practice.⁸

The nutritional evidence considered elsewhere in this book addresses a related but distinct question. PRECISE compared higher and standard enteral protein provision in mechanically ventilated patients, with recovery assessed over six months. Higher protein did not improve recovery and produced a less favourable primary health-utility result. It does not directly refute PROTECTION: the route, treatment duration, patient population and intended therapeutic objective differed. Equally, PROTECTION cannot be used to dismiss the concerns raised by sustained higher protein feeding during critical illness. Both trials argue for specifying exactly which exposure is being proposed and which benefit is expected.^{1,9}

By the end of 2024, the evidence therefore supported a focused conclusion. Brief intravenous amino acid administration around elective cardiac surgery could reduce early creatinine-defined acute kidney injury. The major remaining questions were whether this represented durable protection, which patients derived worthwhile absolute benefit, and whether effects extended to dialysis or recovery beyond hospital discharge. Future

trials should address these outcomes directly and retain the dose and timing detail needed for reproducible practice. A treatment that improves a functional kidney endpoint warrants consideration on that basis, with its broader benefits still to be demonstrated.^{1,2}

Does perioperative intravenous amino acid infusion protect kidney function after elective cardiac surgery?

It reduces creatinine-defined acute kidney injury with the regimen tested in PROTECTION. Benefits in dialysis use, long-term renal recovery and survival remain unproven.

References

1. Landoni G, Monaco F, Ti LK, Baiardo Redaelli M, Bradic N, Comis M, et al.; PROTECTION Study Group. A Randomized Trial of Intravenous Amino Acids for Kidney Protection. *N Engl J Med* 2024;391(8):687–698.
2. Ostermann M, Shaw AD. Amino Acid Infusion to Protect Kidney Function after Cardiac Surgery. *N Engl J Med* 2024;391(8):759–760.
3. Landoni G, Monaco F, Ti LK, Baiardo Redaelli M, Bradic N, Comis M, et al. Supplementary appendix to the PROTECTION trial. *N Engl J Med* 2024;391(8):687–698. Published supplementary methods and results.
4. Bellomo R, Monaco F, Paternoster G. Intravenous Amino Acids and Kidney Protection. Reply. *N Engl J Med* 2024;391(20):1965–1966.
5. Doig GS, Simpson F, Bellomo R, Heighes PT, Sweetman EA, Chesher D, et al. Intravenous amino acid therapy for kidney function in critically ill patients: a randomized controlled trial. *Intensive Care Med* 2015;41(7):1197–1208.
6. Pu H, Doig GS, Heighes PT, Allingstrup MJ, Wang A, Brereton J, et al. Intravenous amino acid therapy for kidney protection in cardiac surgery patients: A pilot randomized controlled trial. *J Thorac Cardiovasc Surg* 2019;157(6):2356–2366.
7. Kazawa M, Kabata D, Yoshida H, Minami K, Maeda T, Yoshitani K, et al. Amino acids to prevent cardiac surgery-associated acute kidney injury: a randomized controlled trial. *JA Clin Rep* 2024;10:19.

8. Pruna A, Losiggio R, Landoni G, Kotani Y, Baiardo Redaelli M, Veneziano M, et al.; PROTECTION Study Group. Amino Acid Infusion for Perioperative Functional Renal Protection: A Meta-analysis. *J Cardiothorac Vasc Anesth* 2024;38(12):3076–3085.
9. Bels JLM, Thiessen S, van Gassel RJJ, Beishuizen A, De Bie Dekker A, Fraipont V, et al.; PRECISE study team. 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.

Possible Reporting Inconsistencies. The main text gives 4,415 screened patients, whereas supplementary Figure S1 gives 4,931. In Table 4, the noninvasive ventilation counts (518/1,759 versus 615/1,752) imply a crude relative risk of approximately 0.84, inconsistent with the printed 0.94 and its associated interval. Supplementary Table S8 lists 595 placebo stage 1 cases (28.3%), which exceeds the total of 561 acute kidney injury cases and does not match the percentage. These discrepancies remain unresolved; the primary intention-to-treat event counts are consistent.

REVIVAL

Pickkers P, Angus DC, Bass K, Bellomo R, van den Berg E, Bernholz J, et al.; REVIVAL investigators. Phase-3 trial of recombinant human alkaline phosphatase for patients with sepsis-associated acute kidney injury (REVIVAL). *Intensive Care Med* 2024;50(1):68–78.

Introduction

Sepsis-associated acute kidney injury combines a major clinical burden with a difficult therapeutic target. Inflammation, altered microcirculation, metabolic adaptation and underlying kidney disease can contribute in different proportions. A drug that modifies one pathway may improve a renal measure without changing the wider course of sepsis. Conversely, a potentially useful kidney treatment may have little measurable effect on mortality when death is also driven by infection severity and failure of other organs. Choosing and interpreting the trial endpoint is therefore central to evaluating a proposed treatment.¹

Alkaline phosphatase offers a biologically plausible approach. The enzyme removes phosphate groups from inflammatory substrates, including bacterial endotoxin and extracellular adenosine triphosphate. These reactions can reduce inflammatory signalling and promote formation of adenosine. Ilofotase alfa is a recombinant human alkaline phosphatase combining properties of intestinal and placental isoenzymes to achieve biological activity and stability. It was developed as a pharmacological treatment to accompany conventional sepsis care, rather than as a replacement for antimicrobial treatment, source control or organ support.¹

Earlier studies had suggested renal effects, and the phase 2 STOP-AKI trial generated an exploratory survival signal despite failing its primary short-term renal endpoint. That combination made a confirmatory trial necessary. A favourable secondary result can justify another experiment, but it cannot reliably predict that experiment's outcome.^{1,2}

REVIVAL asked whether a three-day course of ilofotase alfa improved 28-day survival in patients with recent sepsis-associated acute kidney injury requiring vasopressors. The study stopped for futility. Its subsequent report also presented potentially favourable kidney outcomes, making it a useful examination of two separate questions: whether the mortality hypothesis succeeded, and what the remaining renal findings can reasonably support.¹

Synopsis

REVIVAL was an international, multicentre, double-blind, randomised, placebo-controlled phase 3 trial with a group-sequential design. Recruitment occurred between November 2020 and July 2022 across 107 sites in North America, Europe, Japan and Oceania. Adults required recent acute kidney injury associated with sepsis and ongoing vasopressor support. Eligibility included an initial requirement for at least 0.1 µg/kg/min of noradrenaline or equivalent for one hour despite fluid resuscitation. Subsequent vasopressor doses could be lower, but treatment had to start within 72 hours of commencing continuous vasopressors.^{1,3}

Acute kidney injury could be identified through creatinine or urine-output criteria. Study medication had to begin within 24 hours of kidney injury when sepsis preceded it, or within 48 hours when kidney injury preceded vasopressor treatment. Patients with a current estimated glomerular filtration rate above 60 mL/min/1.73 m² were excluded. Other exclusions included a predominantly non-septic cause of injury, prolonged pre-existing kidney replacement therapy, advanced liver disease, severe immunosuppression, planned extracorporeal membrane oxygenation and specified limits on organ support. These criteria selected an early but already substantially ill population.^{1,3}

Three cohorts were included. The main cohort had a pre-injury reference estimated glomerular filtration rate of at least 45 mL/min/1.73 m² without suspected or proven COVID-19. A smaller chronic kidney disease cohort allowed reference values from 25 to below 45 mL/min/1.73 m² at selected sites. A third cohort enrolled patients with COVID-19-associated sepsis and kidney injury. Results for the combined population included all three, with cohort classification also informed by documented pre-injury renal function.^{1,3}

An independent statistician generated a block randomisation schedule, stratified by site and modified Sequential Organ Failure Assessment score, excluding the neurological component. Allocation was implemented through an interactive voice or web system. Ilofotase alfa was given at 1.6 mg/kg, equivalent to 1,000 units/kg, capped at 192 mg above 120 kg. Patients received a one-hour infusion daily for three consecutive days or matching placebo, alongside usual care. Clinical teams remained blinded and were not permitted to measure serum alkaline phosphatase before day 14.¹

The primary endpoint was all-cause mortality at 28 days. A maximum main-cohort sample of 1,400 was intended to provide approximately 85% power for an eight-percentage-point absolute reduction from an anticipated 35% placebo mortality. The first interim assessment included 411 treated main-cohort patients. Mortality was

29.3% with ilofotase alfa and 25.7% with placebo, and the reported predictive probability of eventual success was 1.2%. The monitoring committee recommended stopping for futility; no safety concern prompted that recommendation.¹

The main renal secondary endpoint was presented in two forms. MAKE90A included death by day 90, an estimated filtration-rate fall exceeding 25% at day 90, kidney replacement therapy through day 28 or at day 90, or rehospitalisation after discharge. MAKE90B included death by day 90, a filtration-rate fall exceeding 25% at both days 28 and 90, or kidney replacement therapy at day 90. The second definition therefore placed greater emphasis on persistent dysfunction and omitted rehospitalisation and earlier transient dialysis.¹

Other secondary outcomes included days alive without organ support, days alive and discharged from intensive care through day 28, and survival through day 90. Because the primary endpoint was not met, subsequent tests were exploratory without adjustment for multiple comparisons. The report primarily presented observed combined-population event proportions, nominal one-sided P values and sensitivity models accounting for the cohorts. Supplementary tables also provided Kaplan–Meier estimates, which should be distinguished from simple observed proportions.^{1,3}

The flow diagram records 655 randomised patients, of whom 650 started treatment. Three received the opposite treatment, giving safety groups of 330 ilofotase alfa and 320 placebo recipients. One placebo recipient lacked efficacy data, leaving 649 patients in the combined efficacy analysis, presented according to treatment received: 330 versus 319. This included 567 main-cohort, 49 chronic kidney disease and 33 COVID-19 patients. Median age was 70 years, approximately 36% were women and three quarters were mechanically ventilated. About 42% had stage 1 acute kidney injury and 27% stage 3.^{1,3}

At day 28, 92 of 330 ilofotase alfa recipients and 89 of 319 placebo recipients had died, both 27.9%. The absolute difference was –0.02 percentage points (95% confidence interval –6.92 to 6.88), with a nominal one-sided P value of 0.50. Day-90 mortality was 33.9% versus 34.8%. Adjustment for cohort did not establish a mortality benefit. In the main cohort alone, observed day-28 mortality was 27.8% versus 24.7%; the small additional cohorts contributed differently to the combined estimate.^{1,3}

MAKE90A occurred in 187 of 330 patients (56.7%) versus 206 of 319 (64.6%), difference –7.91 percentage points (95% confidence interval –15.40 to –0.42; nominal one-sided P=0.019). The largest component difference involved kidney replacement therapy through day 28 or at day 90: 28.2% versus 36.4%. MAKE90B occurred in 120 patients (36.4%) versus 128 (40.1%), difference –3.76 percentage points (–11.24 to 3.71; nominal

one-sided $P=0.162$). Persistent dialysis at day 90 was uncommon, with zero versus two patients recorded.^{1,3}

Median days alive without organ support were 17 versus 14, and days alive and discharged from intensive care were 15 versus 10; neither comparison established a difference. Serious adverse events occurred in 43.3% and 44.1% of the safety populations. A post hoc model suggested greater MAKE90A benefit at lower pre-injury filtration rates, with an interaction P value of approximately 0.024. This was an exploratory finding rather than a validated treatment-selection rule.^{1,3}

Critique

The principal conclusion is clear: REVIVAL did not establish that ilofotase alfa improves survival in this setting. The final combined mortality proportions were virtually identical, and the interim main-cohort result did not suggest a developing survival advantage. Blinding, central allocation and an objective primary endpoint support the credibility of that result. Stopping for futility was consistent with the low reported probability of meeting the trial's mortality objective. It should not be reframed as premature termination of a trial that was already succeeding on its primary question.¹

Futility nevertheless differs from proof of no effect. The final confidence interval for combined day-28 mortality allowed absolute benefit or harm of almost seven percentage points, and recruitment stopped well below the planned maximum. A smaller survival effect remained unresolved. The original assumption of an eight-point absolute reduction was ambitious, while placebo mortality was lower than anticipated. These features constrain precision, but they do not supply evidence that continuing recruitment would have revealed benefit. The appropriate conclusion is failure to demonstrate improved survival, with residual uncertainty about more modest effects.¹

The analysis populations require particular care. Randomisation initially allocated 331 patients to ilofotase alfa and 324 to placebo, whereas the displayed combined efficacy groups were 330 and 319 after excluding untreated patients, accounting for crossovers and omitting one patient without efficacy data. Treatment-received analyses do not retain all the protections of the original randomised assignment. Only three crossovers were recorded, so it would be excessive to assume they explain the findings, but the main combined results should not be described as an unrestricted intention-to-treat comparison of everyone randomised.^{1,3}

Pooling the cohorts adds another interpretative layer. The main cohort contributed most patients, while the chronic kidney disease and COVID-19 cohorts were small. In the

COVID-19 subgroup, observed mortality was five of 20 versus 11 of 13, an extreme imbalance based on very few patients. Such a result cannot establish efficacy in COVID-19, but it influences a pooled proportion. Cohort-adjusted models were useful sensitivity analyses; they could not create precise independent estimates for these small populations or turn the combined analysis into the original main-cohort mortality test.^{1,3}

MAKE90A is clinically interesting because it includes more than a laboratory threshold. Avoiding kidney replacement therapy or another hospital admission matters to patients. However, a favourable exploratory composite following a failed primary endpoint remains vulnerable to chance, particularly when several outcomes, definitions and subgroup models are examined. Its nominal confidence interval only narrowly excluded zero. The result is a credible reason to study renal outcomes further, but not a confirmatory demonstration of a broad clinical benefit. Calculating a practice-facing number needed to treat from it would obscure that uncertainty.¹

The distinction between MAKE90A and MAKE90B is substantive. A patient who receives transient dialysis early and then recovers can satisfy the first definition but not the second, unless another component occurs. Similarly, rehospitalisation appears only in MAKE90A, while the persistent filtration criterion is stricter in MAKE90B. The favourable broader composite therefore cannot be translated directly into prevention of lasting kidney failure. At day 90, dialysis dependence was rare, and the more persistent composite had an interval compatible with benefit and no meaningful difference.^{1,3}

Component interpretation also depends on ascertainment and timing. Death, dialysis, filtration change and rehospitalisation differ in severity, frequency and susceptibility to clinical decisions. Their percentages overlap and cannot simply be added. The published RRT component covers treatment through day 28 or status at day 90; describing this as every dialysis episode throughout 90 days would broaden the specified window. Usual clinical criteria guided dialysis, but practices across 107 sites could still vary. Blinding reduces the risk that treatment knowledge influenced those decisions without making local practice identical.¹

Handling of incomplete follow-up is another reason to examine the supplementary results. The MAKE90A subgroup model assumed that patients without a confirmed composite event did not have one. Absence of confirmation is not always equivalent to complete event-free follow-up. Kaplan–Meier estimates for the overall composite were slightly different from observed proportions, and the supplementary MAKE90A difference had a confidence interval crossing zero. These analyses address follow-up differently; none alone proves that missing observations caused the apparent benefit, but the variation reinforces the limited certainty of the secondary finding.^{1,3}

The pre-injury kidney-function interaction is plausible but remains exploratory. Patients with reduced reserve could have greater potential absolute benefit, yet fitting a treatment interaction after observing the data increases the opportunity to discover a favourable pattern. The reported curve does not establish a clinical threshold near 90 mL/min/1.73 m², nor does it demonstrate benefit in advanced kidney disease outside eligibility. Estimating pre-injury function also depends on the available reference creatinine. A future trial could test an explicitly defined renal-risk population prospectively; current subgroup modelling cannot substitute for that trial.^{1,3}

The safety findings are reassuring within the observed exposure, especially the similar serious-event proportions. The nominal reduction in patients with any adverse event does not establish that ilofotase alfa is broadly protective against complications, and 650 treated patients cannot exclude uncommon harms. AM-Pharma funded and coordinated the study, with external organisations conducting operational and analytical work. This involvement does not invalidate the results, but makes transparent analytical reporting particularly valuable. The March 2024 correction repaired investigator attribution and the contributor roster; it did not announce changes to efficacy outcomes. Clinical adoption would still require stronger evidence of a worthwhile benefit.^{1,4}

Body of Evidence

Early alkaline phosphatase studies were primarily mechanistic. Heemskerk and colleagues studied 36 patients with severe sepsis or septic shock in a double-blind phase 2 trial. Treatment was associated with improved creatinine measurements and changes in markers related to nitric oxide production and tubular injury. Detailed mechanistic measurements involved a subgroup of 15. These findings helped connect enzyme activity with a possible renal effect, but the small population and physiological outcomes could not establish improved survival or a reliable reduction in dialysis. The historical value of the study lies in supporting biological plausibility and the decision to pursue larger experiments.⁵

A separate 2012 trial by Pickkers and colleagues enrolled 36 patients with sepsis-associated acute kidney injury. A 48-hour alkaline phosphatase infusion improved a combined assessment of renal function over 28 days, with a greater rise in endogenous creatinine clearance. Reductions in dialysis requirements and duration were not statistically established. Inflammatory and urinary injury markers also favoured treatment. This was another encouraging small study, but its mixture of renal measurements, limited sample and different preparation meant that it could support drug development more readily than a claim of established patient benefit.⁶

STOP-AKI then evaluated recombinant human alkaline phosphatase in a phase 2 dose-finding programme enrolling 301 patients. The selected 1.6 mg/kg regimen did not improve the primary measure of creatinine clearance over the first week. Later renal and survival findings were exploratory and motivated further study. REVIVAL supplied the more direct test of the mortality hypothesis, and that test was unsuccessful. This sequence illustrates why an attractive secondary finding from a phase 2 trial needs confirmation, even when there is a coherent mechanism and supportive preclinical work.^{1,2}

A February 2024 secondary analysis by Bruse and colleagues explored clinical phenotypes within REVIVAL itself. After exclusions and removal of outliers, 570 patients remained. A more severely ill phenotype appeared to have a favourable MAKE90 treatment association. However, variable selection considered observed treatment effects, and the final model was chosen partly for the strongest separation in outcomes. This creates substantial scope for optimistic estimates when selection and evaluation use the same dataset. It was not an independent replication, and its renal composite also differed from MAKE90A by omitting rehospitalisation. The analysis offers a hypothesis for prospective validation, not a ready-made prescribing algorithm.⁷

The renal literature elsewhere in this book reinforces the importance of defining what a treatment has achieved. PROTECTION concerned a short amino acid infusion around elective cardiac surgery, before or during a predictable insult. REVIVAL treated established sepsis-associated injury with a different drug and a mortality primary endpoint. The success of one functional renal prevention strategy cannot be used to validate another treatment, and a failed mortality trial does not answer every narrower kidney question. Comparisons are most useful when they clarify these differences rather than pool fundamentally different therapeutic claims.^{1,8}

At the end of 2024, ilofotase alfa therefore remained a promising biological intervention without confirmed survival efficacy. Repeated renal signals justified further investigation, but the strongest support for them still came from exploratory analyses and outcomes whose definitions varied between studies. A convincing next trial would need a clearly defined population, a clinically meaningful renal endpoint, consistent handling of death and missing follow-up, and enough patients to assess the proposed benefit. REVIVAL narrows the claim that can be made now while providing useful information for designing that experiment.¹

Does ilofotase alfa improve outcomes in sepsis-associated acute kidney injury?

REVIVAL did not demonstrate improved survival. Its favourable exploratory kidney composite warrants further study but does not establish a benefit sufficient to support routine use.

References

1. Pickkers P, Angus DC, Bass K, Bellomo R, van den Berg E, Bernholz J, et al.; REVIVAL investigators. Phase-3 trial of recombinant human alkaline phosphatase for patients with sepsis-associated acute kidney injury (REVIVAL). *Intensive Care Med* 2024;50(1):68–78.
2. Pickkers P, Mehta RL, Murray PT, Joannidis M, Molitoris BA, Kellum JA, et al.; STOP-AKI Investigators. Effect of Human Recombinant Alkaline Phosphatase on 7-Day Creatinine Clearance in Patients With Sepsis-Associated Acute Kidney Injury: A Randomized Clinical Trial. *JAMA* 2018;320(19):1998–2009.
3. Pickkers P, Angus DC, Bass K, Bellomo R, van den Berg E, Bernholz J, et al. Supplementary information to the REVIVAL trial. *Intensive Care Med* 2024;50(1):68–78. Published tables, figures and eligibility criteria.
4. Pickkers P, Angus DC, Bass K, Bellomo R, van den Berg E, Bernholz J, et al.; REVIVAL investigators. Correction: Phase-3 trial of recombinant human alkaline phosphatase for patients with sepsis-associated acute kidney injury (REVIVAL). *Intensive Care Med* 2024;50(4):614–615.
5. Heemskerk S, Masereeuw R, Moesker O, Bouw MPWJM, van der Hoeven JG, Peters WHM, et al.; APSEP Study Group. Alkaline phosphatase treatment improves renal function in severe sepsis or septic shock patients. *Crit Care Med* 2009;37(2):417–423, e1.

6. Pickkers P, Heemskerk S, Schouten J, Laterre PF, Vincent JL, Beishuizen A, et al. Alkaline phosphatase for treatment of sepsis-induced acute kidney injury: a prospective randomized double-blind placebo-controlled trial. *Crit Care* 2012;16:R14.
7. Bruse N, Pardali K, Kraan M, Kox M, Pickkers P; REVIVAL investigators. Phenotype-specific therapeutic efficacy of ilofotase alfa in patients with sepsis-associated acute kidney injury. *Crit Care* 2024;28:50.
8. Landoni G, Monaco F, Ti LK, Baiardo Redaelli M, Bradic N, Comis M, et al.; PROTECTION Study Group. A Randomized Trial of Intravenous Amino Acids for Kidney Protection. *N Engl J Med* 2024;391(8):687–698.

Possible Reporting Inconsistencies. The abstract labels MAKE90B percentages of 37.4% and 42.8% as observed, whereas supplementary Table 4 identifies them as Kaplan–Meier estimates; the observed proportions in Table 2 are 36.4% and 40.1%. The day-90 placebo dialysis entry in Table 2 is printed as ‘0 (2/319)’; supplementary Table 4 gives 0.6% (2/319). The March 2024 correction concerns contributor attribution and does not resolve these presentation issues.

SIRAKI02

Pérez-Fernández X, Ulsamer A, Cámara-Rosell M, Sbraga F, Boza-Hernández E, Moret-Ruiz E, et al.; SIRAKI02 Study Group. Extracorporeal Blood Purification and Acute Kidney Injury in Cardiac Surgery: The SIRAKI02 Randomized Clinical Trial. JAMA 2024;332(17):1446–1454.

Introduction

Cardiac surgery creates an unusually predictable opportunity to prevent acute kidney injury. The insult occurs at a known time, patients can be assessed beforehand, and cardiopulmonary bypass provides access to an extracorporeal circuit. However, postoperative kidney dysfunction has multiple causes, including altered perfusion, venous congestion, haemolysis and inflammation. An intervention aimed at one mechanism may change early creatinine or urine output without preventing severe or persistent injury. A successful prevention trial therefore needs interpretation beyond whether its primary endpoint crosses a statistical threshold.¹

Adsorptive blood purification seeks to remove circulating inflammatory mediators during bypass. The oXiris membrane combines blood purification with adsorption of cytokines and endotoxin. Connecting it to the bypass circuit permits treatment during the period of surgical inflammatory activation. This approach differs from commencing kidney replacement therapy after organ failure has developed: its intended purpose is prevention, and its potential value depends on improving outcomes enough to justify the additional equipment and circuit management.¹

Previous haemoadsorption studies had shown that biochemical activity did not reliably translate into clinical benefit. Their devices, populations and outcomes varied, making it inappropriate to assume either a universal beneficial effect or universal futility. A trial of oXiris in a defined cardiac surgical population was needed to establish what this particular treatment strategy could achieve.^{1,2}

SIRAKI02 evaluated intraoperative blood purification in patients undergoing non-emergency cardiac surgery with expected prolonged bypass. Its favourable primary result was a reduction in acute kidney injury within seven days. The central questions are how securely that difference can be attributed to the assigned strategy, whether it reflects a clinically durable benefit, and how far a study from two centres can support wider adoption.¹

Synopsis

SIRAKI02 was a randomised trial conducted at two high-volume university hospitals in Spain. Recruitment ran from June 2016 to November 2021, with follow-up ending in February 2022. Adults undergoing non-emergency cardiac surgery were eligible when cardiopulmonary bypass was expected to exceed 90 minutes. Patients with stage 4 or 5 chronic kidney disease, defined by an estimated glomerular filtration rate below 30 mL/min/1.73 m², were excluded. Other exclusions included recent kidney replacement therapy, transplantation, immunosuppressive treatment and conditions associated with limited life expectancy.^{1,3}

Of 3,815 potentially eligible patients, 1,156 underwent eligibility assessment. Staffing, timing and pandemic-related constraints prevented assessment of many others, while consent and clinical exclusions further reduced recruitment. Ultimately, 343 patients were randomised: 169 to extracorporeal blood purification and 174 to standard care. The flow diagram records six duplicate envelope openings for patients already randomised, which were resolved by retaining a single assignment for each patient. All 343 contributed to the primary outcome.¹

A computer-generated sequence used variable blocks of 10–30, with allocation stratified by centre. Sealed envelopes were opened by a research assistant in the operating room. Patients and postoperative clinicians assessing outcomes were blinded, but the operating team could see whether the device was being used. The report's description as double-blind therefore applies to particular participants in the trial, rather than to everyone delivering perioperative treatment.¹

The intervention connected an oXiris filter to the bypass circuit using a Prismaflex machine in slow continuous ultrafiltration mode. Treatment was confined to the intraoperative bypass period, with mean blood flow of approximately 276 mL/min. The perfusionist could remove fluid when considered necessary. Control patients received standard bypass care, with a conventional polysulfone filter available for ultrafiltration when required. There was no sham adsorptive circuit, and the study did not test continued postoperative purification in intensive care.^{1,3}

The primary endpoint was acute kidney injury within seven days, classified using Kidney Disease: Improving Global Outcomes creatinine and urine-output criteria. Baseline creatinine was measured before surgery, with frequent postoperative measurements during the first day and daily measurements thereafter through day seven. Urine output was generally available during the first four days, while patients remained catheterised in intensive care. The immediate postoperative hours were excluded from urine-based classification because mannitol could influence diuresis.¹

The intended sample was 340, based on an anticipated 25% control event rate and a ten-percentage-point absolute reduction, with an allowance for losses. The primary analysis compared groups without adjustment; additional regression analyses examined the result after accounting for selected covariates. Secondary outcomes included injury severity, kidney replacement therapy, survival, organ-support-free days, lengths of stay and later creatinine measurements. Cytokine changes and exploratory subgroup and sensitivity analyses were also reported.^{1,3}

Patients were about 69 years old on average, and 35% were women. Approximately one fifth had pre-existing stage 1–3 chronic kidney disease, while diabetes affected 33% of intervention patients and 28% of controls. Baseline creatinine averaged approximately 1.0 mg/dL and left ventricular ejection fraction approximately 58%. Combined procedures and double-valve operations were common. Mean bypass duration was 132 minutes with purification and 127 minutes with standard care.¹

Acute kidney injury occurred in 48 of 169 intervention patients (28.4%) and 69 of 174 controls (39.7%). The absolute reduction, expressed as control minus intervention risk, was 11.25 percentage points (95% confidence interval 1.30–21.21; $P=0.03$). The adjusted absolute reduction was 10.42 percentage points (2.34–18.49). The unadjusted point estimate corresponds to approximately nine patients treated to prevent one episode meeting the trial definition, although the interval around that estimate was wide.¹

Stage 2 or 3 injury occurred in 22 intervention patients and 34 controls, an absolute difference of 6.52 percentage points (95% confidence interval –1.25 to 14.30; $P=0.08$). Kidney replacement therapy was required in three versus six patients, without an established difference. Peak creatinine averaged 1.26 versus 1.39 mg/dL, and oliguria was recorded in 13% versus 20%; neither individual comparison established a statistically significant difference. The separately printed control stage 3 count is internally inconsistent and is addressed below.^{1,3}

Among 96 patients whose injury began within 48 hours, 39 were in the intervention group and 57 in the control group. Injury resolving within 48 hours was classified as transient, occurring in 23 of 39 and 35 of 57, respectively; the remainder were classified as persistent. These percentages describe the course among patients who developed early injury. They are not whole-group estimates of a randomised effect on persistent kidney injury.^{1,3}

Survival at 90 days was approximately 95% versus 96%, without an established difference. Median intensive care stay was three days in both groups, and hospital stay 13 days in both. Organ-support-free days were counted until discharge or day 28,

whichever came first, and did not differ clearly. Day-90 creatinine was available for 127 intervention patients and 125 controls, with no established between-group difference.¹

Ultrafiltration was used in 41 intervention patients (24%) and 21 controls (12%). Among those receiving it, median volume removed was lower with the intervention, at 480 versus 700 mL. A post hoc analysis excluding these patients still favoured purification for acute kidney injury, at 37 of 128 versus 63 of 153. Cytokine measurements, available in 205 patients, suggested attenuation of some intraoperative inflammatory changes. Reported perioperative complications and platelet changes did not establish a consistent safety difference.^{1,3}

Critique

SIRAKI02 provides a positive randomised result for a clinically recognisable early renal endpoint. Complete primary-outcome ascertainment among the randomised patients, an explicit injury definition and blinded postoperative assessment are meaningful strengths. The absolute difference was substantial, and its direction persisted in the reported adjusted analysis. It would be unreasonable to dismiss this solely because dialysis and mortality were not significantly different: with only nine dialysis events and 16 deaths by day 90, the study could not precisely evaluate those outcomes.¹

The uncertainty around the primary estimate still matters. Its confidence interval ranged from a modest reduction of approximately one percentage point to a reduction exceeding 20 points. The control event rate was considerably higher than the 25% assumed in planning. A number needed to treat near nine describes the observed setting and point estimate, rather than a stable expectation for every cardiac surgical service. Replication is particularly valuable when an initially favourable effect is large, the sample modest and the intervention adds practical complexity.¹

Blinding was incomplete at the point where the intervention could influence care. The operating team could alter fluid removal, fluid administration or other management in response to knowing the assignment. The observed difference in ultrafiltration use shows that this was more than an abstract possibility. It does not prove that ultrafiltration caused the renal benefit, but it prevents a clean attribution of the entire treatment effect to adsorption of inflammatory mediators. The randomised comparison evaluates the implemented purification strategy, including any associated changes in care.¹

Excluding patients who received ultrafiltration does not remove this uncertainty. Ultrafiltration was a decision made after randomisation and may have reflected evolving

physiology, clinical judgement or the assigned device. Conditioning the analysis on that decision selects different subsets from the original groups. The continued favourable association is informative, but the remaining 128 and 153 patients are not newly randomised comparable populations. This sensitivity analysis cannot establish that adsorption independently accounts for the primary finding, nor can it substitute for a study that standardises fluid-removal decisions.^{1,3}

The adjusted analysis also needs restraint. The supplementary model included postoperative Sequential Organ Failure Assessment score alongside age, body mass index, chronic kidney disease and bypass duration. The renal component was removed from the score, reducing direct overlap with the outcome, but postoperative non-renal dysfunction could still be affected by treatment. Adjusting for it can alter the causal question rather than simply correct baseline imbalance. The unadjusted randomised comparison should therefore remain the main estimate. The paper and supplement also label the reported relative measure differently, reinforcing the value of focusing on the directly observed absolute risks.^{1,3}

The primary endpoint is relevant but heterogeneous. A transient change in creatinine, an episode of oliguria and dialysis-requiring injury do not have identical consequences. Creatinine is influenced by fluid balance, while urine ascertainment became less complete after intensive care discharge. Differential fluid management could affect both physiology and how thresholds were crossed. Nevertheless, the published data do not justify claiming that most injuries were diagnosed solely through urine output: the report does not provide the necessary cross-classification of creatinine and urine criteria.^{1,3}

Similarly, the early-injury analysis should not be simplified into proof that purification prevents only transient injury. The recovery percentages are conditional on developing early injury, an event that treatment might itself influence. Similar proportions recovering among those selected patients do not imply the same whole-population risk of persistent injury. Moreover, the 96 early cases do not include every primary-outcome event. The most defensible conclusion is that the trial reduced seven-day injury overall, while the severity, duration and longer-term consequences of that reduction remain less certain.^{1,3}

The secondary outcomes do not establish durable renal protection. The combined stage 2–3 estimate favoured treatment but remained imprecise, and dialysis events were scarce. Approximately one quarter of participants lacked day-90 creatinine measurements. Similar values among those measured cannot exclude a difference among those missing, and comparisons restricted to survivors require additional caution.

Likewise, similar survival and length of stay should be described as absence of demonstrated benefit, rather than proof of equivalence or proof that preventing early injury has no value.¹

Mechanistic findings support further investigation without proving mediation. Cytokine sampling covered only about 60% of the randomised population, several markers were assessed repeatedly, and results were not uniformly favourable. The reported confidence interval for the between-group difference in interleukin-8 change crossed zero, despite stronger language elsewhere in the article. A measurable difference in circulating concentrations does not show that the kidney endpoint changed because of that difference. Nor does a nonselective adsorptive membrane establish that every removed substance was harmful.^{1,3}

Subgroup findings are particularly vulnerable to overinterpretation. Apparent benefits in diabetes, chronic kidney disease or reduced ventricular function were based on limited numbers. The displayed P values tested effects within subgroups, rather than demonstrating a treatment interaction between them. Some ventricular-function categories were very small. These analyses cannot identify a validated subgroup for selective clinical use. A plausible higher-risk population is an appropriate subject for another trial, but statistical significance in one subgroup and its absence in another does not establish different treatment effects.¹

Generalisability and safety remain bounded by the recruitment setting. Only 343 of 3,815 potentially eligible patients entered the trial, at two centres familiar with the technique, and advanced chronic kidney disease and emergency operations were excluded. Similar observed platelet changes and complications are reassuring but do not exclude uncommon device harms. Some supplementary complication data were also incomplete. The study received Spanish public funding and a Baxter grant; the authors reported no funder role in conduct, analysis or publication. Overall, SIRAIO2 supports confirmatory investigation more strongly than routine implementation.^{1,3}

Body of Evidence

Early trials demonstrate why technical feasibility and clinical efficacy must be considered separately. Bernardi and colleagues randomised 46 cardiac surgical patients to intraoperative CytoSorb haemoadsorption or control. Thirty-seven remained in the principal analysed cohort, with further losses for cytokine analyses. The study did not demonstrate a reduction in its principal inflammatory outcomes or a clear improvement in the perioperative course. Its small size and exclusions limited inference, but it established that adding an adsorber to bypass did not automatically suppress the postoperative inflammatory response. This was evidence about a different device, not a direct test of oXiris.⁴

REMOVE provided a larger randomised assessment in a different, highly inflammatory setting: surgery for infective endocarditis. It assigned 288 patients to intraoperative CytoSorb treatment or standard care. The primary comparison of postoperative change in Sequential Organ Failure Assessment score was negative, and 30-day mortality was approximately 21% versus 22%. Some circulating cytokines were reduced, but clinically important outcomes did not improve demonstrably. The distinction between biochemical activity and organ benefit is directly relevant to interpreting SIRAki02, although differences in infection, baseline risk and adsorptive technology prevent treating REMOVE as a replication.²

RECCAS, published in December 2024, reinforced that distinction. Thirty-eight older patients undergoing elective cardiac surgery were randomised to intraoperative CytoSorb treatment or standard care. The primary interleukin-6 concentration on intensive care admission did not differ. Measurements across the adsorber suggested removal of several mediators during treatment, but this did not translate into consistently lower postoperative concentrations or improved clinical outcomes. With only 19 patients per group, the trial could not establish effects on uncommon events. It illustrates how removal across a device and the patient's subsequent systemic response can give different answers.⁵

The CORONARY kidney substudy provides a useful perspective on early renal endpoints. Among 2,932 patients, off-pump coronary bypass reduced postoperative acute kidney injury compared with on-pump surgery, from 20.8% to 17.5%. However, loss of kidney function at one year was not reduced, occurring in 17.1% versus 15.3%. This does not show that preventing early injury is unimportant or that SIRAki02 cannot produce lasting benefit. It shows why such benefit must be measured directly: an intervention can improve a short-term renal definition without establishing preservation of longer-term function.⁶

PROTECTION approached the same broad prevention problem through intravenous amino acids rather than adsorption. In more than 3,500 cardiac surgical patients, amino acids reduced creatinine-defined acute kidney injury, with a favourable secondary stage 3 result. Effects on dialysis and survival remained uncertain, and later renal follow-up was incomplete. Its larger sample strengthens the evidence for that specific intervention but does not validate extracorporeal purification. The two trials differ in mechanism, injury definition and treatment burden, so their absolute risk reductions should not be compared as though they were competing treatments tested in the same patients.⁷

By the end of 2024, SIRAKI02 had added an important positive renal result to an otherwise inconsistent haemoadsorption literature. Its contribution is specific: an intraoperative oXiris strategy reduced early acute kidney injury in selected non-emergency cardiac surgical patients. It did not establish a class effect for blood purification, a mortality benefit, a dialysis benefit or a reliable subgroup-selection rule. The accompanying editorial emphasised the need to connect early changes with outcomes that persist and matter to patients.^{1,8}

A decisive next study would reproduce the intervention across more centres, standardise fluid-removal decisions, preserve allocation in the analysis and collect renal outcomes after discharge with minimal missingness. Severe or persistent injury, dialysis, survival and treatment burden would help determine whether the early effect warrants the additional circuit. Until those questions are resolved, the most useful response to SIRAKI02 is to recognise the favourable primary finding while keeping the clinical claim within the limits of the experiment.¹

Should intraoperative oXiris purification be used routinely to prevent kidney injury after cardiac surgery?

SIRAKI02 reduced early acute kidney injury, but did not establish less dialysis or lasting renal benefit. Further confirmation is needed before routine use.

References

1. Pérez-Fernández X, Ulsamer A, Cámara-Rosell M, Sbraga F, Boza-Hernández E, Moret-Ruíz E, et al.; SIRAKI02 Study Group. Extracorporeal Blood Purification and Acute Kidney Injury in Cardiac Surgery: The SIRAKI02 Randomized Clinical Trial. *JAMA* 2024;332(17):1446–1454.

2. Diab M, Lehmann T, Bothe W, Akhyari P, Platzer S, Wendt D, et al.; REMOVE Trial Investigators. Cytokine Hemoadsorption During Cardiac Surgery Versus Standard Surgical Care for Infective Endocarditis (REMOVE): Results From a Multicenter Randomized Controlled Trial. *Circulation* 2022;145(13):959–968.
3. Pérez-Fernández X, Ulsamer A, Cámara-Rosell M, Sbraga F, Boza-Hernández E, Moret-Ruiz E, et al. Supplementary online content to the SIRAKI02 randomized clinical trial. *JAMA* 2024;332(17):1446–1454. Published figures and tables.
4. Bernardi MH, Rinoesl H, Dragosits K, Ristl R, Hoffelner F, Opfermann P, et al. Effect of hemoadsorption during cardiopulmonary bypass surgery—a blinded, randomized, controlled pilot study using a novel adsorbent. *Crit Care* 2016;20:96.
5. Hohn A, Malewicz-Oeck NM, Buchwald D, Annecke T, Zahn PK, Baumann A. REmoval of cytokines during CARDiac surgery (RECCAS): a randomised controlled trial. *Crit Care* 2024;28:406.
6. Garg AX, Devereaux PJ, Yusuf S, Cuerden MS, Parikh CR, Coca SG, et al.; CORONARY Investigators. Kidney Function After Off-Pump or On-Pump Coronary Artery Bypass Graft Surgery: A Randomized Clinical Trial. *JAMA* 2014;311(21):2191–2198.
7. Landoni G, Monaco F, Ti LK, Baiardo Redaelli M, Bradic N, Comis M, et al.; PROTECTION Study Group. A Randomized Trial of Intravenous Amino Acids for Kidney Protection. *N Engl J Med* 2024;391(8):687–698.
8. Kwong YD, Liu KD. Impact of Adsorptive Blood Purification on Kidney Outcomes. *JAMA* 2024;332(17):1430–1431.

Possible Reporting Inconsistencies. Table 2 prints control stage 3 acute kidney injury as ‘1 (6)’, which is inconsistent with both the percentage and the reported stage totals; the combined stage 2–3 count is used here without silently correcting the separate entry. The main text labels the adjusted relative effect an odds ratio, whereas supplementary eTable 6 labels it a risk ratio. Later correction: on 21 February 2025, JAMA corrected the control-group confidence interval in the visual abstract; that notice does not resolve these separate reporting issues.



Haematology Trials

WOMAN-2

WOMAN-2 Trial Collaborators. The effect of tranexamic acid on postpartum bleeding in women with moderate and severe anaemia (WOMAN-2): an international, randomised, double-blind, placebo-controlled trial. Lancet 2024;404(10463):1645–1656.

Introduction

Postpartum haemorrhage can become life-threatening before there is time to organise definitive treatment. Women who are already anaemic enter childbirth with reduced capacity to tolerate blood loss, and severe anaemia is associated with a greater risk of bleeding and serious complications. Preventing haemorrhage in this population is therefore especially attractive. However, a treatment that reduces the consequences of established bleeding does not necessarily prevent bleeding from beginning, particularly when delivery itself produces the initiating injury.¹

Tranexamic acid inhibits fibrinolysis and helps preserve formed fibrin. The earlier WOMAN trial supported its early use to reduce death from bleeding in women with postpartum haemorrhage. This raised the possibility that administering the drug before haemorrhage was diagnosed could offer additional protection. Such prophylaxis would expose many more women, including those who would not otherwise bleed substantially, making the balance between benefit, treatment burden and uncommon harm different from emergency treatment.^{1,2}

Anaemia provided both a clinical reason to target prevention and a proposed biological rationale. Changes in circulatory physiology and clot properties might make bleeding more severe or harder to control. Yet these mechanisms were not established as treatment-responsive features in every anaemic woman. The relationship between anaemia and obstetric complications also reflects other factors, including access to antenatal care and underlying illness. A large randomised trial was needed rather than extrapolation from risk alone.¹

WOMAN-2 tested one practical strategy: 1 g of intravenous tranexamic acid shortly after vaginal birth, administered over ten minutes after cord clamping. Its negative result is consequential because recruitment, blinding and follow-up were strong. The challenge is to interpret that result precisely, recognising what it establishes about this regimen while distinguishing unanswered questions about earlier administration, severe bleeding and treatment after haemorrhage has already developed.¹

Synopsis

WOMAN-2 was an international, randomised, double-blind, placebo-controlled trial conducted at 34 hospitals in Nigeria, Pakistan, Tanzania and Zambia. Women of any age in active labour were screened using a point-of-care haemoglobin measurement. Eligibility required a concentration below 100 g/L and no existing indication or contraindication to tranexamic acid. Moderate anaemia was defined as 70–99 g/L and severe anaemia as below 70 g/L. Women planning vaginal birth were approached during labour, with consent procedures adapted to their clinical circumstances.¹

Final eligibility was confirmed around delivery of the baby's anterior shoulder and before cutting or clamping the cord. Caesarean delivery or haemorrhage developing before cord clamping could make a woman ineligible. Participants received either 1 g of tranexamic acid in 10 mL or matching saline placebo, administered intravenously at approximately 1 mL/min. Treatment was to begin as soon as possible, no later than 15 minutes after cord clamping. Randomisation was considered to occur when administration started.¹

An independent information technology specialist and statistician generated the allocation codes. Treatment was assigned by taking the lowest numbered pack from a box of 20 visually identical packs. Each contained two ampoules, a syringe and a needle. The central team monitored use of the allocation sequence, and independent chemical testing checked pack contents. Participants, clinicians, research teams and coordinating staff remained masked. Emergency unblinding was available but was not required for any participant.¹

The primary outcome was a clinical diagnosis of postpartum haemorrhage within 24 hours of randomisation. This could be an estimated blood loss above 500 mL or any blood loss sufficient to compromise haemodynamic stability. Clinical teams estimated loss using records, blood-soaked pads and supporting pictograms. They judged haemodynamic compromise from findings such as low blood pressure, tachycardia and reduced urine output. Calibrated blood-collection drapes were not used.¹

Secondary measures included estimated blood loss, haemoglobin corrected for transfusion, interventions for bleeding, transfusion, a death-or-near-miss composite, organ dysfunction, functional recovery and safety. The near-miss definition included severe bleeding, selected surgical interventions, failure to form clots, major transfusion and cardiovascular or renal dysfunction. Many clinical and recovery outcomes were collected until discharge, death or day 42, whichever came first. After discharge, women received contact details for reporting medical events through day 42; this was not equivalent to active assessment of every outcome at six weeks.^{1,3}

The published final sample-size calculation targeted 15,000 women, providing approximately 85% power to detect a 15% relative reduction with a 9% placebo event rate. Analyses retained assigned treatment regardless of whether the full dose was administered. Sensitivity analyses adjusted for baseline risk factors and excluded women with continuing antepartum bleeding. The report examined heterogeneity by anaemia severity, antepartum bleeding, birth canal trauma, pain control and modelled baseline bleeding risk, using a stringent reported threshold of $P < 0.001$ for strong subgroup evidence.^{1,3}

Between August 2019 and September 2023, 16,586 women were assessed and 15,068 randomised: 7,580 to tranexamic acid and 7,488 to placebo. One woman in each group lacked primary-outcome data, leaving 15,066 in the main analysis. Seven received only one ampoule because the other was broken, and 15 had recorded eligibility or treatment deviations; these women remained in the analysis. The combination of near-complete follow-up and retention of deviations preserved the intended comparison well.¹

Mean age was approximately 27 years and mean haemoglobin 83 g/L. Around 86% had moderate and 14% severe anaemia. Mean gestation was 37 weeks, and most pregnancies were singleton. Nearly three quarters of participants were recruited in Pakistan. Prophylactic oxytocin was used in 99.9% of both groups. Median time from birth to cord clamping was two minutes, and from birth to starting study treatment five minutes.^{1,3}

Clinically diagnosed postpartum haemorrhage occurred in 530 of 7,579 women receiving tranexamic acid (7.0%) and 497 of 7,487 receiving placebo (6.6%): risk ratio 1.05 (95% confidence interval 0.94–1.19; $P = 0.39$). Adjustment for baseline factors and exclusion of continuing antepartum bleeding did not reveal benefit. Uterine atony was the most frequently assigned cause, followed by tears and retained placental tissue. No clear reduction emerged for the broader distribution of estimated blood loss.¹

Mean estimated blood loss was 309.8 versus 310.8 mL, a difference of -0.95 mL (95% confidence interval -7.10 to 5.21). Mean corrected haemoglobin was 82.2 versus 82.1 g/L. The reported death-or-near-miss composite occurred in 122 women (1.6%) versus 137 (1.8%), risk ratio 0.88 (0.69–1.12). Any blood product was given to 25.4% versus 25.3%, and hysterectomy occurred in nine versus ten women. There were nine in-hospital maternal deaths in each group, with four versus three attributed to bleeding.¹

No vascular occlusive events were reported in either group, and none of the separately recorded adverse events was considered treatment-related. No clear differences appeared in quality of life, anaemia symptoms, walking distance, organ dysfunction or

higher-level care. Median time to discharge was approximately 13 hours. Non-trial tranexamic acid was used after randomisation in 5.8% versus 5.4%, reflecting the availability of treatment when clinically required.^{1,3}

Timing analyses showed that 358 of 1,024 women with haemorrhage and available timing data were diagnosed before the ten-minute injection had finished. Median time from starting treatment to diagnosis was 18.5 minutes overall and 13 minutes among women with severe anaemia. In those whose treatment started within five minutes of birth, the risk ratio was 0.98 (95% confidence interval 0.85–1.13). The antepartum-haemorrhage subgroup had a higher event rate with tranexamic acid, with an interaction P value of 0.044. This did not meet the report's threshold for strong evidence of heterogeneity.^{1,3}

Critique

The trial gives strong evidence against the proposed moderate preventive effect on its primary endpoint. The lower confidence limit of 0.94 excludes the 15% relative reduction used in the final sample-size calculation. Although placebo haemorrhage incidence was lower than anticipated, the observed interval is more informative than a general statement that the study was underpowered. A small reduction remained compatible with the data, as did an increase, but a clinically persuasive preventive benefit was not demonstrated.¹

Internal validity is a major strength. Matching placebo, independent allocation coding, chemical checks, central monitoring and almost complete outcome collection make systematic allocation-related bias less likely. The inclusion of women with incomplete treatment or recorded deviations is preferable to excluding inconvenient cases after randomisation. The unequal final group sizes do not themselves indicate a flaw in randomisation. With only two missing primary outcomes, attrition cannot plausibly account for a large hidden benefit.¹

The clinical primary endpoint has advantages and limitations. Anaemic women can deteriorate after a volume of blood loss that would be better tolerated by a woman with a normal haemoglobin concentration, so a fixed volume threshold alone may miss clinically important haemorrhage. Conversely, visual estimates and clinical judgement are imperfect, and local thresholds for recognising instability can vary. Blinding reduces differential judgement, while training improves consistency; neither removes measurement error. The trial most directly answers whether this strategy reduced recognised haemorrhage within the clinical system studied.¹

Supportive measurements make it difficult to attribute the whole result to a poorly measured primary outcome. Estimated blood loss, corrected haemoglobin, transfusion and clinical interventions did not show a coherent benefit hidden behind the diagnostic label. The post hoc calculated-blood-loss analysis also failed to establish benefit, with a risk ratio of 0.96 (95% confidence interval 0.87–1.06). However, this calculation used a threshold of at least 1,000 mL and assumptions about blood volume and transfusion correction. It was a related alternative outcome, not an objective remeasurement of the identical clinical endpoint.^{1,3}

Timing is the strongest biological question raised by the trial. Although the drug was described as prophylactic, treatment began after birth, when bleeding from the placental bed or birth canal could already be occurring. A diagnosis made before the injection finished might represent an event that the regimen had little opportunity to prevent. This limits inference about earlier administration. It does not invalidate the randomised comparison: delayed pharmacological opportunity is part of how this particular strategy performed in practice.¹

The timing data do not prove that an earlier regimen would succeed. Starting within five minutes was not associated with a clear benefit, although even this remained after delivery and before completion of a ten-minute injection. Actual treatment time was not randomised and could reflect clinical circumstances, staffing or difficulty administering the drug. Comparisons across timing categories therefore cannot establish an optimal minute or justify faster injection. A trial assigning an earlier strategy would need to assess both benefit and the consequences of different maternal and fetal exposure.^{1,3}

The preventive question must also remain separate from rescue treatment. Both groups could receive non-trial tranexamic acid when indicated, with similar use. Such treatment could lessen differences in downstream outcomes, but excluding recipients would select women according to events after randomisation and would distort the policy comparison. WOMAN-2 tested adding routine early prophylaxis to care that already included uterotonics and access to haemorrhage treatment. It did not test withholding tranexamic acid from women who subsequently developed established postpartum haemorrhage.¹

Severe clinical outcomes were much less precise than the primary endpoint. Only 18 in-hospital deaths occurred, including seven attributed to bleeding. The death-or-near-miss confidence interval allowed a worthwhile reduction as well as no benefit. Thus, the study cannot establish mortality equivalence, nor can a numerically favourable composite overcome the negative primary result. Transfusion occurred in a

quarter of participants, substantially more often than diagnosed haemorrhage; in this anaemic population it cannot be interpreted solely as a marker of major new blood loss.¹

Safety follow-up requires similar calibration. No recorded thromboses is reassuring, but it does not establish zero thrombotic risk. Many women left hospital within a day, and subsequent adverse-event reporting depended on women contacting the study team. This differs from systematic investigation for events after discharge. The supplementary organ-dysfunction table also recorded nine seizures with tranexamic acid and three with placebo. The small numbers and multiple comparisons do not establish causation, but they make an unqualified statement that there were no potentially relevant safety imbalances inappropriate.^{1,3}

The antepartum-bleeding subgroup is a reason for caution, not a definitive harm claim. Its risk ratio was 1.44, with a confidence interval excluding one, but the interaction P value was 0.044 amid several subgroup tests. The investigators appropriately required stronger evidence before treating heterogeneity as established. They did not record sufficient coagulation measurements or the presumed cause of antepartum bleeding to prove their proposed explanation. In particular, the suggestion of late antifibrinolytic treatment during evolving coagulopathy remains a hypothesis rather than a demonstrated mechanism.¹

Nor does baseline risk identify a proven group that benefits. Severe anaemia was associated with more haemorrhage, but treatment effects did not convincingly differ by severity. The prognostic model's low, intermediate and high categories were derived within this dataset, using risk-score cut points among women who bled. They were not equal thirds of the whole population or an externally validated prescribing tool. High underlying risk increases the potential absolute value of an effective treatment; it does not make an ineffective regimen effective.^{1,3}

The setting is highly relevant to a major burden of maternal illness, with substantial recruitment in South Asia and Africa. Generalisation nevertheless requires attention to the predominance of Pakistan, hospital birth, near-universal oxytocin, vaginal delivery and the exclusion of women with an existing treatment indication. The trial does not answer prophylaxis before caesarean delivery or management where uterotonics are unavailable. Funding came from the Gates Foundation and Wellcome Trust, with no reported funder role in the analysis or publication. Its negative result is a valuable limit on routine treatment expansion.¹

Body of Evidence

The first WOMAN trial addressed treatment of established postpartum haemorrhage. It randomised more than 20,000 women and found fewer deaths from bleeding with tranexamic acid, particularly when given within three hours of birth. Its primary composite of death from any cause or hysterectomy was not reduced, and an all-cause mortality benefit was not established. The therapeutic inference rests on the bleeding-specific result and its timing, not success on every outcome. WOMAN-2 asked a different question, so its failure to prevent a new clinical diagnosis does not overturn that evidence for early treatment of recognised haemorrhage.²

TRAAP examined prophylaxis after vaginal birth alongside oxytocin. Of 4,079 women randomised, 3,891 delivered vaginally. The primary collector-bag measure of blood loss of at least 500 mL occurred in 8.1% with tranexamic acid and 9.8% with placebo, risk ratio 0.83 (95% confidence interval 0.68–1.01). The primary result did not establish benefit, although clinician-assessed haemorrhage and additional uterotonic use were reduced. These secondary findings were encouraging, but the study did not show that routine prophylaxis reliably prevented its principal measured bleeding outcome.⁴

TRAAP2 produced a positive primary result after caesarean delivery. Among women with outcome data, calculated blood loss above 1,000 mL or red-cell transfusion by day two occurred in 26.7% versus 31.6%, adjusted risk ratio 0.84 (95% confidence interval 0.75–0.94). However, gravimetrically measured loss, clinically assessed haemorrhage and transfusion considered separately did not improve demonstrably. This is an important positive prevention result, but its calculated composite and surgical population differ from WOMAN-2. It supports careful endpoint-specific interpretation rather than a universal claim about prophylactic tranexamic acid.⁵

The subsequent US Maternal–Fetal Medicine Units Network trial randomised 11,000 women undergoing caesarean delivery to tranexamic acid or placebo after cord clamping. Maternal death or transfusion occurred in 3.6% versus 4.3%, adjusted risk ratio 0.89 (95.26% confidence interval 0.74–1.07). The primary outcome was not significantly reduced. Some bleeding-related interventions were less frequent, while postpartum infectious complications were more frequent. This again distinguishes modest secondary changes from confirmation of the principal clinical benefit and cautions against treating all prophylaxis studies as uniformly positive.⁶

Ker and colleagues' 2024 meta-analysis combined 54,404 women from five large trials, with individual data from four and aggregate data from one. Life-threatening bleeding occurred in 0.65% versus 0.85%, pooled odds ratio 0.77 (95% confidence interval 0.63–0.93). Thromboembolic events were not clearly increased. The synthesis included

WOMAN-2 and combined treatment and prevention studies, so it was neither independent replication nor a direct demonstration that this anaemia-based prophylactic regimen worked. Lack of detected subgroup heterogeneity cannot establish identical effects across all clinical settings.⁷

E-MOTIVE placed tranexamic acid within a broader response to haemorrhage. Eighty hospitals were randomised to usual care or an intervention combining calibrated blood-loss detection with uterine massage, uterotonics, tranexamic acid, fluids, examination and escalation. The composite of severe haemorrhage, laparotomy for bleeding or death from bleeding occurred in 1.6% versus 4.3%, risk ratio 0.40 (95% confidence interval 0.32–0.50). This supports timely recognition and coordinated treatment, but cannot isolate the contribution of tranexamic acid from the rest of the bundle.⁸

The accompanying WOMAN-2 editorial highlighted the difference between large, carefully conducted trials and more optimistic estimates from smaller studies. It also emphasised timing and selection for preventive treatment. Those questions remained open at the end of 2024. The direct evidence from WOMAN-2 was narrower and firm: routinely starting this dose after cord clamping did not prevent clinically diagnosed haemorrhage in women with moderate or severe anaemia. Effective antenatal care, uterotonic prophylaxis, prompt recognition and treatment of bleeding remained essential; an unconfirmed extension of tranexamic acid prophylaxis could not replace them.^{1,9}

Does tranexamic acid after vaginal birth prevent postpartum haemorrhage in women with anaemia?

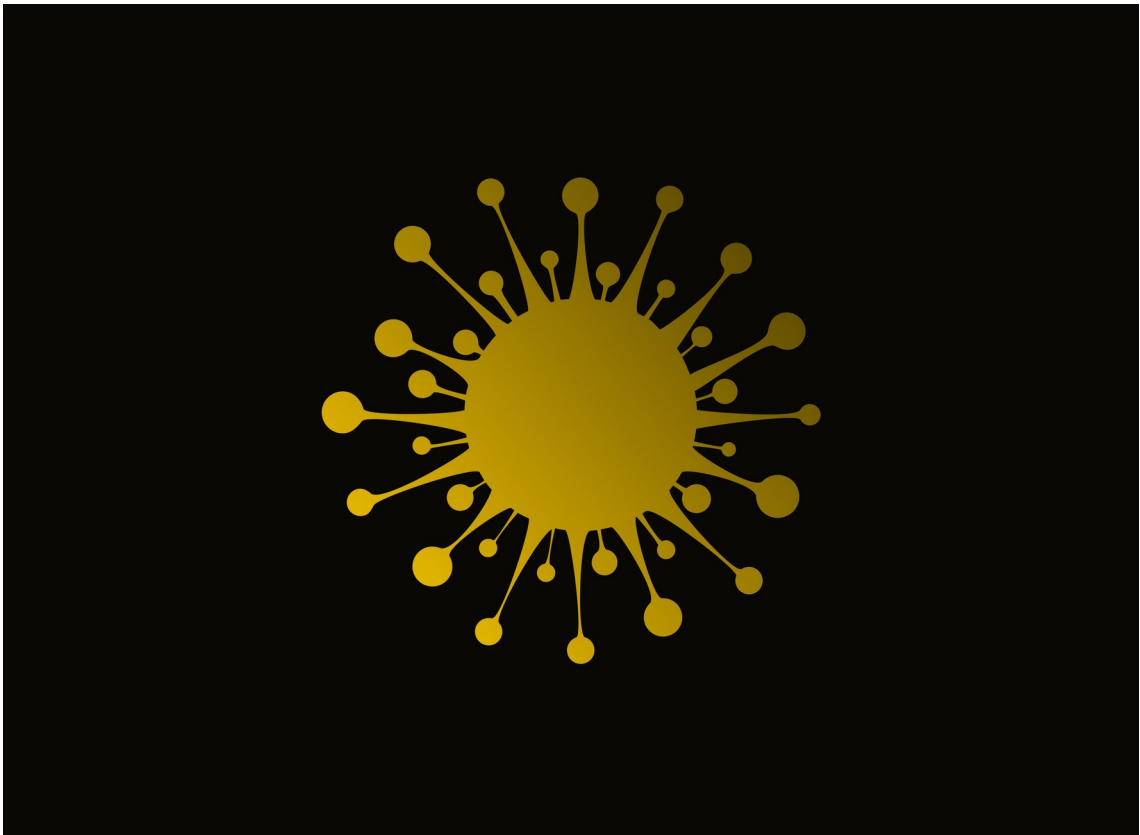
WOMAN-2 found no preventive benefit from 1 g started after cord clamping. This does not overturn evidence supporting early treatment of established postpartum haemorrhage.

References

1. WOMAN-2 Trial Collaborators. The effect of tranexamic acid on postpartum bleeding in women with moderate and severe anaemia (WOMAN-2): an international, randomised, double-blind, placebo-controlled trial. *Lancet* 2024;404(10463):1645–1656.

2. WOMAN Trial Collaborators. Effect of early tranexamic acid administration on mortality, hysterectomy, and other morbidities in women with post-partum haemorrhage (WOMAN): an international, randomised, double-blind, placebo-controlled trial. *Lancet* 2017;389(10084):2105–2116.
3. WOMAN-2 Trial Collaborators. Supplementary appendix to the WOMAN-2 trial. *Lancet* 2024;404(10463):1645–1656. Published supplementary methods, tables and figures.
4. Sentilhes L, Winer N, Azria E, Sénat MV, Le Ray C, Vardon D, et al.; Groupe de Recherche en Obstétrique et Gynécologie. Tranexamic Acid for the Prevention of Blood Loss after Vaginal Delivery. *N Engl J Med* 2018;379(8):731–742.
5. Sentilhes L, Sénat MV, Le Lous M, Winer N, Rozenberg P, Kayem G, et al.; Groupe de Recherche en Obstétrique et Gynécologie. Tranexamic Acid for the Prevention of Blood Loss after Cesarean Delivery. *N Engl J Med* 2021;384(17):1623–1634.
6. Pacheco LD, Clifton RG, Saade GR, Weiner SJ, Parry S, Thorp JM Jr, et al.; NICHD Maternal–Fetal Medicine Units Network. Tranexamic Acid to Prevent Obstetrical Hemorrhage after Cesarean Delivery. *N Engl J Med* 2023;388(15):1365–1375.
7. Ker K, Sentilhes L, Shakur-Still H, Madar H, Deneux-Tharaux C, Saade G, et al.; Anti-fibrinolytics Trialists Collaborators Obstetric Group. Tranexamic acid for postpartum bleeding: a systematic review and individual patient data meta-analysis of randomised controlled trials. *Lancet* 2024;404(10463):1657–1667.
8. Gallos I, Devall A, Martin J, Middleton L, Beeson L, Galadanci H, et al. Randomized Trial of Early Detection and Treatment of Postpartum Hemorrhage. *N Engl J Med* 2023;389(1):11–21.
9. Mol BWJ, Flanagan M, Li W. Prevention and treatment of postpartum haemorrhage: progress by learning what works and what does not work. *Lancet* 2024;404(10463):1616–1618.

Possible Reporting Inconsistencies. Table 2 labels the death-or-near-miss composite as assessed at 24 hours, while the Methods place this outcome among those assessed at death, discharge or 42 days, whichever came first. The event counts are reported here without treating the composite as a fully observed 42-day outcome.



Sepsis Trials

BLING III

Dulhunty JM, Brett SJ, De Waele JJ, Rajbhandari D, Billot L, Cotta MO, et al.; BLING III Study Investigators. Continuous vs Intermittent β -Lactam Antibiotic Infusions in Critically Ill Patients With Sepsis: The BLING III Randomized Clinical Trial. JAMA 2024;332(8):629–637.

Introduction

Choosing an appropriate antibiotic is only part of treating sepsis. The prescribed drug must also reach its target at a concentration that remains effective between doses. For β -lactam antibiotics, bacterial killing depends particularly on the time during which the unbound concentration exceeds the organism's minimum inhibitory concentration. Giving the same daily quantity continuously offers a plausible way of increasing that exposure without increasing the total dose. The question is whether this pharmacological advantage translates into better survival.^{1,2}

Critical illness makes the answer less straightforward than the infusion chart suggests. Changes in distribution volume and renal clearance can produce very different concentrations from the same prescription. A continuous infusion may help when clearance is rapid, but it cannot compensate for choosing an inactive antibiotic, inadequate source control or an insufficient daily dose. Conversely, sustained exposure could increase toxicity when elimination is impaired. An effective administration strategy must therefore be judged within the broader treatment of infection.^{1,3}

By 2021, the Surviving Sepsis Campaign suggested prolonged β -lactam maintenance infusions after an initial bolus, but the recommendation was weak. Earlier randomised trials had established feasibility and pharmacokinetic advantages without conclusively resolving mortality. Practice consequently varied, with some units routinely using continuous administration and others retaining short intermittent infusions.^{1,4}

BLING III addressed this uncertainty on an unprecedented scale. It compared continuous with 30-minute intermittent administration of piperacillin–tazobactam or meropenem in more than 7,000 critically ill adults. Published alongside an updated meta-analysis, it generated two related questions: what did this individual trial establish, and how should its findings change the interpretation of the accumulated evidence? Those questions require separate answers.^{1,2}

Synopsis

BLING III was an international, open-label, phase 3 randomised trial conducted in 104 intensive care units across Australia, Belgium, France, Malaysia, New Zealand, Sweden and the United Kingdom. Recruitment ran from March 2018 to January 2023, with 90-day follow-up completed in April 2023. The report appeared online on 12 June 2024. Patients, treating teams and those administering the antibiotics knew the allocation, although the report describes blinded outcome adjudication.¹

Eligible adults had a documented or strongly suspected infection, had commenced piperacillin–tazobactam or meropenem within the preceding 24 hours, and had recent organ dysfunction. Qualifying dysfunction included hypotension, sustained vasopressor use, respiratory support or substantial creatinine elevation. The treating clinician had to consider either administration method appropriate and anticipate a continuing ICU requirement. Important exclusions were kidney replacement therapy at randomisation, pregnancy, relevant antibiotic allergy, previous enrolment, imminent inevitable death or a decision not to provide advanced life support for the next 48 hours.^{1,5}

Allocation was centralised through a secure web system, using a minimisation algorithm stratified by site. Patients were assigned equally to continuous or intermittent infusion. Clinicians selected the antibiotic and total daily dose, with subsequent changes allowed according to the patient's condition. This compared administration methods at otherwise comparable daily doses; it did not impose one fixed prescription on every patient. Piperacillin–tazobactam accounted for approximately four-fifths of treatment before randomisation, with meropenem comprising most of the remainder.^{1,5}

Continuous treatment delivered the daily prescription across 24 hours, whereas intermittent doses were infused over 30 minutes. A loading dose preceded continuous administration; an appropriate dose already given could fulfil that requirement. Switching between the two study antibiotics retained the assigned administration method and required another loading dose in the continuous group. The published supplementary methods specify assigned treatment in the ICU through day 14, with local administration practice after ICU discharge or thereafter. Therapeutic drug monitoring was not permitted during randomised treatment.^{1,5}

Continuous administration did not necessarily mean using a single bag for an entire day. Preparation followed local policy, product information and stability requirements. Meropenem commonly required several freshly prepared bags or syringes over 24 hours. Other antibiotics, decisions to change or stop treatment, and accompanying sepsis care remained clinical decisions. The median reported duration on study treatment was 5.8

days with continuous and 5.7 days with intermittent infusion. Daily study antibiotic doses were similar between groups.^{1,5}

The primary outcome was all-cause mortality at 90 days. Secondary outcomes were clinical cure by day 14, new acquisition or infection with a multiresistant organism or *Clostridioides difficile* within 14 days, and ICU and hospital mortality. Clinical cure required completion of the antibiotic course by day 14 without restarting antibiotics for the same infection within 48 hours. Escalation or de-escalation for the same infection remained part of that course. Hospital discharge within 14 days generally counted as cure, subject to reassessment after an early readmission; death during treatment did not.^{1,5}

The investigators calculated that 7,000 participants would provide 90% power to detect a 3.5 percentage-point absolute mortality reduction from 27.5%, allowing for withdrawals and missing follow-up. The primary model used logistic regression with treatment as a fixed effect and site as a random effect. An additional analysis adjusted for sex, APACHE II score, admission source and antibiotic used before randomisation. Secondary and tertiary outcome families received Holm–Bonferroni multiplicity correction. Missing values were not imputed.¹

Of 29,042 patients screened, 7,202 were randomised. Consent or approval for data use was unavailable for 171, leaving 3,498 assigned continuous infusion and 3,533 assigned intermittent infusion in the modified intention-to-treat cohort. Their mean age was 59 years, approximately 35% were women and the mean APACHE II score was about 20. Around 70% had received vasopressors or inotropes in the preceding 24 hours. The primary infection site was pulmonary in 59.5%, intra-abdominal in 13.0%, bloodstream in 8.0% and urinary in 5.4%.¹

At 90 days, 864 of 3,474 patients with known outcomes in the continuous group had died, compared with 939 of 3,507 in the intermittent group: 24.9% versus 26.8%. The reported absolute difference was –1.9 percentage points (95% CI –4.9 to 1.1), with an odds ratio of 0.91 (95% CI 0.81–1.01; $P=0.08$). Fifty participants lacked day-90 status. The additional adjusted analysis produced an odds ratio of 0.89 (95% CI 0.79–0.99; $P=0.04$). No convincing interaction emerged across the five reported subgroup comparisons.¹

Clinical cure occurred in 1,930 of 3,467 patients receiving continuous infusion and 1,744 of 3,491 receiving intermittent infusion, 55.7% versus 50.0%. The absolute difference was 5.7 percentage points (95% CI 2.4–9.1), with an odds ratio of 1.26 (95% CI 1.15–1.38). New multiresistant-organism or *C. difficile* events occurred in 7.2% versus 7.5%. ICU mortality was 17.1% versus 18.4%, and hospital mortality 23.3% versus 25.0%; neither

comparison established a difference. No tertiary measure of days alive and free of ICU, hospital, ventilation or kidney replacement therapy was significant after multiplicity adjustment.¹

There were ten reported adverse events with continuous and six with intermittent infusion. One continuous-infusion patient developed severe encephalopathy followed by aspiration, cardiac arrest and death; this was considered possibly related to meropenem administration. Recorded deviations affected roughly one-third of patients, although many concerned individual delays or doses. The trial received public and institutional funding. The George Institute sponsored the study and conducted its statistical analysis; the listed funders were reported to have no role in the trial's conduct or reporting.^{1,5}

Critique

BLING III provides a substantially more precise estimate than its predecessors, but its primary result remains uncertain. The mortality point estimate favours continuous infusion, and the interval includes a benefit that would matter greatly for an inexpensive change in routine care. It also includes no benefit and some harm. Describing the trial as proving equivalence would therefore be incorrect. Equally, converting the observed approximately two percentage-point reduction into a definitive number needed to treat of 50 would suppress the uncertainty surrounding that estimate.¹

The adjusted analysis strengthens the suggestion of benefit without replacing the primary analysis. Its covariates are clinically reasonable, and adjustment can improve precision even when randomisation has produced balanced groups. Nevertheless, two analyses of the same deaths are not independent replications. The distinction between $P=0.08$ and $P=0.04$ should not dominate the clinical interpretation: their estimates are similar and their uncertainty overlaps substantially. The appropriate conclusion is that mortality results consistently lean towards benefit, while the trial's primary statistical comparison did not establish superiority.¹

Mortality is an important strength. Death at 90 days matters directly to patients and is considerably less vulnerable to observer judgement than clinical improvement or antibiotic discontinuation. Follow-up was also highly complete within the consented cohort. The 171 patients excluded because data use was not authorised are a separate issue from the 50 missing day-90 outcomes. Their exclusion means this was not an unrestricted analysis of everyone randomised, although the proportion was small. Allocation concealment, broad recruitment and good baseline balance otherwise support the comparison's internal validity.¹

Clinical cure is more difficult to interpret. Its explicit definition reduces inconsistency in assigning the label, but cannot remove the influence of unblinded treatment decisions that determine whether the definition is met. Clinicians knew which infusion a patient received when deciding to stop, change or restart antibiotics and when considering discharge. A reproducibly applied rule can therefore remain vulnerable to performance bias. The 5.7 percentage-point improvement is clinically interesting and survived multiplicity correction, but it should not be treated as a direct demonstration of faster bacterial eradication.^{1,5}

The cure definition also combines several pathways. Early hospital discharge generally qualified, while remaining on antibiotics beyond day 14 did not. Escalation and de-escalation could both belong to the same treatment course. Supplementary cessation data show substantial differences in recorded reasons for changing or completing the study antibiotic, without demonstrating why these occurred. These observations warrant caution about attributing the cure effect to a single biological mechanism. They do not invalidate the result, but they make its meaning broader than microbiological resolution alone.⁵

Actual treatment delivery adds another layer. Deviations were common at the patient level: approximately 32% and 34% had at least one recorded event. Incorrect administration methods occurred in 8.1% of continuous and 4.8% of intermittent recipients, whereas pauses or delays longer than an hour were more common with intermittent treatment. These categories should not be mistaken for complete crossover of a third of patients. They show the practical difficulty of maintaining either strategy in an ICU. Analysing by assignment appropriately estimates the effect of adopting the strategy under these conditions; excluding patients afterwards could introduce selection bias.⁵

The trial addresses maintenance administration after initial antibiotic exposure. Most patients had already received intermittent treatment before randomisation, and assigned administration ended when ICU treatment requirements changed. It cannot establish the effect of continuous infusion from the very first antibiotic dose or throughout every subsequent ward day. Initial loading remains essential to the strategy studied. Any attenuation from treatment before randomisation is plausible, but its magnitude cannot be recovered simply by assuming that perfect adherence would have produced a larger survival benefit.^{1,5}

The pharmacological explanation is compelling but incompletely tested in the principal report. Drug concentrations and organism-specific minimum inhibitory concentrations were not incorporated into the main outcome analysis, and therapeutic drug monitoring

was prohibited. Consequently, the trial does not identify which patients actually had inadequate exposure during intermittent treatment or whether achieving a particular concentration mediated the outcomes. Concomitant antibiotics were frequent. Randomisation helps distribute these treatments, but the result remains the effect of changing one component of a broader antimicrobial regimen rather than testing isolated β -lactam monotherapy.^{1,5}

Generalisability is broad within comparable ICUs and narrower outside them. Patients already receiving kidney replacement therapy were excluded, accounting for more than 4,000 screening exclusions. The trial therefore cannot directly determine the optimum strategy in that population. Renal support started after enrolment is a different circumstance: outcomes in those patients cannot substitute for randomising patients already receiving support, because the need for subsequent support may itself be influenced by illness evolution after allocation. Most recruitment occurred in health systems with established critical care infrastructure, and resistance patterns may differ elsewhere. Piperacillin–tazobactam predominated. Absence of a statistically convincing interaction with antibiotic choice does not prove identical effects for meropenem, other β -lactams or infections with different susceptibility profiles.¹

Safety reporting was directed towards events thought possibly related to the assigned administration method, rather than every complication of critical illness. The small number of reports therefore cannot establish equal safety with high precision. The encephalopathy event is relevant because prolonged exposure and impaired clearance may increase neurotoxicity, although a single possibly related event cannot establish causation. Practical adoption still requires appropriate daily dosing, a loading dose, reliable pumps and access, compatible co-infusions, and preparation that respects stability. Continuous infusion changes drug delivery; it does not remove the need to reassess the prescription as organ function changes.^{1,3,5}

Body of Evidence

The early BLING trial enrolled 60 patients and compared continuous with intermittent β -lactam administration using a double-blind design. In the pharmacokinetic analysis, concentrations exceeded the minimum inhibitory concentration in 18 of 22 continuous-infusion patients and six of 21 intermittent-infusion patients. Clinical cure also favoured continuous treatment, but hospital survival and ICU-free days did not differ significantly. This supported the biological rationale and feasibility of a larger trial; the small sample and incomplete pharmacokinetic measurements precluded a reliable mortality conclusion.⁶

BLING II subsequently enrolled 432 patients across 25 ICUs. Its primary outcome, days alive and free of ICU at day 28, was similar: medians of 18 and 20 days with continuous and intermittent treatment. Ninety-day survival was 74.3% versus 72.5%, with a hazard ratio of 0.91 and a wide 95% confidence interval of 0.63–1.31. Clinical cure was also not significantly different. These findings left room for a modest mortality benefit and provided a clear reason to undertake a much larger outcome trial.⁷

The 2023 MERCY trial offered a complementary test restricted to meropenem. In 607 critically ill adults, its double-blind comparison found the composite of death or emergence of extensively drug-resistant or pandrug-resistant bacteria by day 28 in 47% with continuous versus 49% with intermittent administration (relative risk 0.96, 95% CI 0.81–1.13). Ninety-day mortality was 42% in both groups. MERCY therefore supplies no independent confirmation of a survival advantage, although its sample was much smaller than BLING III and its population, antibiotic and primary endpoint differed.⁸

The systematic review published alongside BLING III assembled 18 randomised trials involving 9,108 patients; 17 trials and 9,014 participants contributed to the principal mortality analysis. Using a Bayesian random-effects model with vague priors, the pooled risk ratio was 0.86 (95% credible interval 0.72–0.98). The posterior probability of a mortality reduction of any size was 99.1%. Alternative prior assumptions and frequentist analyses also favoured prolonged infusion. Ten trials judged at low risk of bias supplied approximately 95% of the mortality data.²

This synthesis strengthens the case for prolonged administration, but several details matter. Although labelled 90-day mortality, the pooled outcome used the closest available time point when a study did not report day-90 deaths; some trials contributed ICU or hospital mortality. Seventeen trials evaluated continuous infusion and only one evaluated an extended infusion. The evidence consequently supports prolonged administration principally through studies of continuous treatment. It cannot determine whether a three- or four-hour infusion produces the same benefit as uninterrupted daily administration.²

BLING III contributed approximately 77% of all enrolled participants, calculated from the review's reported totals, but only 37.4% of the weight in its principal Bayesian forest plot. Patient numbers and random-effects statistical weights are not interchangeable. The pooled estimate incorporates smaller trials and between-study variation; it is not merely a recalculation of BLING III's P value. Conversely, citing BLING III and the meta-analysis as two independent large confirmations would count the same patients twice. The posterior probability describes benefit under the model and included evidence, not a 99.1% chance that every patient benefits.²

The review also favoured prolonged administration for ICU mortality, with a risk ratio of 0.84 (95% credible interval 0.70–0.97), and clinical cure, 1.16 (1.07–1.31). Its authors rated mortality evidence high certainty and cure evidence moderate certainty. Definitions of cure varied, and evidence for microbiological cure, adverse events and ICU length of stay remained less secure. No subgroup convincingly identified a population with a different mortality effect. That absence should temper claims of preferential benefit in particular infection or renal-support categories rather than erase their remaining uncertainty.²

Taken together, the evidence available by the end of 2024 supports continuous β -lactam infusion as a reasonable routine approach for comparable adults with sepsis, provided delivery is reliable. BLING III contributes scale, pragmatic relevance and a favourable but inconclusive mortality estimate; the wider randomised evidence increases confidence that some survival benefit is likely. The practical conclusion is stronger than the primary P value alone suggests, but more qualified than a claim that one infusion method has conclusively improved every important outcome.^{1,2,3}

Should β -lactam antibiotics be given by continuous infusion in critically ill adults with sepsis?

Continuous infusion after a loading dose is a reasonable default for suitable patients, supported by the combined evidence, although BLING III alone did not establish a mortality benefit.

References

1. Dulhunty JM, Brett SJ, De Waele JJ, Rajbhandari D, Billot L, Cotta MO, et al.; BLING III Study Investigators. Continuous vs Intermittent β -Lactam Antibiotic Infusions in Critically Ill Patients With Sepsis: The BLING III Randomized Clinical Trial. *JAMA* 2024;332(8):629–637.
2. Abdul-Aziz MH, Hammond NE, Brett SJ, Cotta MO, De Waele JJ, Devaux A, et al. Prolonged vs Intermittent Infusions of β -Lactam Antibiotics in Adults With Sepsis or Septic Shock: A Systematic Review and Meta-Analysis. *JAMA* 2024;332(8):638–648.
3. Wiersinga WJ, van Agtmael MA. Resolving the Dilemma on Continuous vs Intermittent β -Lactam Antibiotics in Sepsis. *JAMA* 2024;332(8):623–625.

4. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med* 2021;47(11):1181–1247.
5. BLING III Study Investigators. Supplement 3 to the BLING III randomised clinical trial. *JAMA* 2024;332(8):629–637. Published supplementary methods, tables and figures.
6. Dulhunty JM, Roberts JA, Davis JS, Webb SAR, Bellomo R, Gomersall C, et al. Continuous infusion of beta-lactam antibiotics in severe sepsis: a multicenter double-blind, randomized controlled trial. *Clin Infect Dis* 2013;56(2):236–244.
7. Dulhunty JM, Roberts JA, Davis JS, Webb SAR, Bellomo R, Gomersall C, et al.; BLING II Investigators for the ANZICS Clinical Trials Group. A Multicenter Randomized Trial of Continuous versus Intermittent β -Lactam Infusion in Severe Sepsis. *Am J Respir Crit Care Med* 2015;192(11):1298–1305.
8. Monti G, Bradić N, Marzaroli M, Konkayev A, Fominskiy E, Kotani Y, et al.; MERCY Investigators. Continuous vs Intermittent Meropenem Administration in Critically Ill Patients With Sepsis: The MERCY Randomized Clinical Trial. *JAMA* 2023;330(2):141–151.

BALANCE

The BALANCE Investigators, for the Canadian Critical Care Trials Group, the Association of Medical Microbiology and Infectious Disease Canada Clinical Research Network, the Australian and New Zealand Intensive Care Society Clinical Trials Group, and the Australasian Society for Infectious Diseases Clinical Research Network. Antibiotic Treatment for 7 versus 14 Days in Patients with Bloodstream Infections. N Engl J Med 2025;392(11):1065–1078. Published online 20 November 2024.

Introduction

A positive blood culture often prompts clinicians to extend an antibiotic course, even when the original infection is improving. The concern is understandable: bacteria in the bloodstream may signal a serious infection, and relapse can be devastating. Yet the presence of bacteraemia does not itself establish how many additional days of treatment are useful. For many patients, the traditional two-week course has rested more securely on convention than on direct comparative evidence.¹

Longer treatment is not cost-free. It increases exposure to drug toxicity, disrupts the microbiome and can select resistant organisms. It also consumes clinical time and may prolong the need for intravenous access. These concerns favour reducing unnecessary treatment, but they cannot justify accepting an important increase in death or recurrent infection. The relevant question is whether a shorter strategy retains clinical effectiveness while meaningfully reducing antibiotic exposure.¹

Earlier randomised trials supported seven rather than fourteen days for selected patients with uncomplicated Gram-negative bacteraemia. Their participants were often clinically stable, infection sources were dominated by the urinary tract, and relatively few patients had critical illness. Composite endpoints and noninferiority margins of ten percentage points also left uncertainty about small but important differences in mortality.¹

BALANCE was designed to address that gap through a large international comparison using 90-day all-cause mortality. More than half its participants were enrolled in intensive care, and the study included a wider range of pathogens and sources than its predecessors. Its result offers a practical basis for shortening treatment, provided the strategy is interpreted alongside the exclusions, actual antibiotic durations and uncertainty within smaller clinical subgroups.¹

Synopsis

The Bacteremia Antibiotic Length Actually Needed for Clinical Effectiveness trial was an investigator-initiated, open-label, randomised noninferiority study conducted in 74 hospitals across Canada, Australia, New Zealand, Saudi Arabia, Israel, Switzerland and the United States. Recruitment, including the pilot phase, extended from October 2014 to May 2023. The study enrolled both ICU and hospital ward patients. Its principal report was first published online on 20 November 2024, before appearing in a 2025 print issue.^{1,2}

Patients were eligible when hospitalised with a blood culture yielding a pathogenic bacterium. Exclusions included severe neutropenia, immunosuppressive treatment for organ or stem-cell transplantation, a prosthetic heart valve or synthetic endovascular graft, and an infection requiring established prolonged treatment. Examples included endocarditis, osteomyelitis, septic arthritis, an undrained abscess or an unremoved infected prosthesis. *Staphylococcus aureus*, *S. lugdunensis* and fungal bloodstream infections were excluded, as were single positive cultures containing a likely contaminant. These criteria were intended to select infections for which either treatment duration was clinically acceptable.^{1,2}

Randomisation used a web system with variable block sizes and stratification by hospital and ICU versus ward location. Patients were assigned equally to seven or fourteen days of adequate antibiotic therapy. Adequacy meant that the isolated organism was susceptible according to the local laboratory. Antibiotic selection, dose, frequency and route remained at the treating team's discretion. Allocation was concealed until day seven of adequate treatment, when the two strategies began to diverge. There was no universal placebo after antibiotics were stopped.¹

Adherence was assessed daily. The published analysis allowed a two-day window either side of the assigned duration: five to nine days for the short strategy and twelve to sixteen days for the longer strategy. This tolerance defined the per-protocol population; it did not change the intended comparison from seven versus fourteen days. Clinicians could extend or recommence antibiotics when they considered this necessary. Treatment therefore represented a planned duration embedded within ongoing clinical reassessment rather than an inflexible prohibition on further antibiotics.¹

The primary outcome was death from any cause by 90 days after the positive index blood culture. The noninferiority margin was an absolute increase of four percentage points. The target sample of 3,626 assumed 22% mortality, 80% power and a one-sided significance level of 2.5%, with allowances for loss to follow-up and interim analyses. The final primary comparison used a two-sided 95.7% confidence interval reflecting the

interim alpha spending. Noninferiority required its upper limit to remain below four percentage points.¹

Additional analyses included the per-protocol population and a modified intention-to-treat population excluding patients who died before day seven of treatment. A secondary model accounted for centre and ICU or ward enrolment. Secondary outcomes covered mortality during admission, recurrent bacteraemia, antibiotic-related adverse events, *C. difficile* infection, newly detected resistant organisms, hospital and ICU stay, organ support and treatment-free days. Death before the relevant assessment horizon was assigned zero treatment-free days. Secondary comparisons were exploratory and their confidence intervals were not adjusted for multiplicity.¹

Among 36,637 patients screened, 13,597 were eligible and 3,631 were randomised. Twenty-three were subsequently excluded because of consent issues or ineligibility, leaving the reported intention-to-treat population of 3,608: 1,814 assigned seven days and 1,794 assigned fourteen days. Median age was 70 years, 53.3% were men and 55.0% were enrolled in ICU. Mechanical ventilation was being used in 21.2%, while approximately 30% received vasopressors at baseline. Canada contributed 76.2% of the cohort.^{1,2}

Three-quarters of bloodstream infections were community-acquired. The urinary tract accounted for 42.2% of sources, the abdomen or hepatobiliary system for 18.8%, the lung for 13.0%, vascular catheters for 6.3% and skin or soft tissue for 5.2%. A source-control procedure was recorded in 44.9%. Monomicrobial Gram-negative infections represented 71.0%, monomicrobial Gram-positive infections 17.3%, and polymicrobial infections 11.7%. *Escherichia coli* was the most common organism, followed by *Klebsiella* species. Baseline patient, infection and treatment characteristics were broadly balanced.¹

Ninety-day status was available for 1,802 short-course and 1,779 long-course patients. Death occurred in 261 and 286, respectively: 14.5% versus 16.1%, an absolute difference of -1.6 percentage points (95.7% CI -4.0 to 0.8). The upper boundary was below the four-point margin, establishing noninferiority for the primary outcome. The per-protocol difference was -2.0 percentage points (95% CI -4.5 to 0.6). Excluding the fourteen early deaths in each group also preserved noninferiority, with a difference of -1.6 points (95% CI -3.9 to 0.7).¹

Actual treatment lasted a median of eight days in the seven-day group and fourteen days in the fourteen-day group. Treatment exceeded the permitted adherence window in

23.1% and 10.7%, respectively; shorter-than-window treatment occurred in 0.8% and 5.8%. Median antibiotic-free days through day 28 were nineteen versus fourteen, a difference of five days (95% CI 4.6–5.4). Hospital stay was a median of ten versus eleven days, while ICU stay was five days in both groups.¹

Recurrent bacteraemia occurred in 47 patients receiving the short strategy and 39 receiving the longer strategy, 2.6% versus 2.2% (difference 0.4 points, 95% CI –0.6 to 1.4). In-hospital mortality was 9.3% versus 10.3%; ICU mortality among patients treated there was 9.0% versus 9.6%. *C. difficile* infection occurred in 1.7% versus 2.0%, and secondary resistant-organism infection or colonisation in 9.5% versus 8.5%. Antibiotic-related adverse events were uncommon. These comparisons did not establish a reduction in drug-related complications or antimicrobial resistance. Public and research funding supported the trial.¹

Critique

BALANCE provides persuasive evidence that a planned seven-day strategy does not cause a large increase in 90-day mortality among eligible patients. Its confidence interval is particularly informative: the upper boundary allows an absolute mortality increase of only 0.8 percentage points, well below the chosen four-point margin. This is more reassuring than merely reporting that a noninferiority test was passed. However, the interval still permits a small disadvantage, and its inclusion of zero means the numerically lower mortality does not establish superiority.¹

The margin deserves scrutiny even though the observed result comfortably clears it. Against a control mortality of 16.1%, four percentage points would represent a substantial proportional increase in deaths. Whether that loss could ever be acceptable in exchange for less antibiotic exposure is a clinical judgement, not a property bestowed by the statistical design. In this trial, the much narrower observed interval reduces dependence on accepting the full margin. It is the precision of the result, together with the reduction in exposure, that makes the finding useful.¹

Nonadherence is the principal threat to interpreting the trial as a comparison of exact durations. Almost a quarter of short-course patients received more than nine days, and the group median was eight. Bringing treatment durations closer together can favour noninferiority even when a perfectly enforced short course would perform differently. The achieved six-day separation in medians and five additional antibiotic-free days nevertheless show that the strategies remained meaningfully distinct. BALANCE supports starting with a shorter plan and reassessing; it does not prove that every

included patient would have done equally well if antibiotics had been stopped precisely after seven days.¹

The per-protocol result is reassuring, but it cannot fully repair this limitation. Patients who adhered were selected partly by events after randomisation. Those with persistent infection, new hospital-acquired infection or a complicated recovery may have been more likely to receive longer courses and be excluded from that analysis. Comparing adherent patients therefore sacrifices some protection from randomisation. Agreement between the primary and per-protocol analyses strengthens confidence, while leaving the effect of rigidly enforced seven-day treatment uncertain.¹

Mortality is an objective, patient-important endpoint, and follow-up exceeded 99% in both analysed groups. The trial also concealed allocation until the planned treatment divergence, limiting opportunities for earlier management to differ because clinicians knew the eventual duration. After day seven, decisions to investigate fever, restart antibiotics, discharge a patient or attribute an adverse event remained susceptible to knowledge of allocation. Blinded resistance adjudication and outcome analysis help with interpretation, but cannot undo differences in the clinical decisions that generated the observations.¹

Including deaths before the strategies diverged could theoretically dilute a treatment effect. Here only fourteen patients in each group died before day seven, and removing them barely changed the estimate. That analysis resolves an important concern without suggesting that early deaths should routinely be excluded from randomised trials. The 23 post-randomisation exclusions from the reported intention-to-treat cohort and the subsequent 27 missing outcomes should also be distinguished. Both numbers were small, but the label intention-to-treat does not mean every originally randomised patient contributed an observed day-90 result.¹

The ICU representation is a major advance. Among patients enrolled there, 90-day deaths were 180 of 996 with shorter and 181 of 986 with longer treatment. The risk difference was -0.3 percentage points (95% CI -3.7 to 3.1), within the study's four-point margin. Yet ICU enrolment is not synonymous with ongoing shock when treatment stops. At day seven, 28.3% of the overall cohort remained in ICU, while 21.8% had gone home. The study includes substantial critical illness without establishing that all persistently unstable patients should have antibiotics stopped at the same moment.^{1,2}

Other subgroups are less precise. Confidence intervals in patients receiving vasopressors, those with higher APACHE II scores, Gram-positive infections and several non-urinary sources extended beyond the noninferiority margin. This does not show that

fourteen days is superior in those groups; it shows that the subgroup estimates cannot independently exclude clinically relevant harm. The overall randomised result remains important, but the phrase broadly consistent should not be translated into a separate proof of noninferiority for every organism, source and severity category.¹

Selection also limits extrapolation. Only 26.7% of eligible patients were randomised, with treating-physician refusal the largest specified reason for non-enrolment. Clinicians may have been especially reluctant to include infections they considered difficult to treat. The trial explicitly excluded severe immunocompromise and several situations in which prolonged therapy or management of an infected device is central. A source-control procedure being recorded in fewer than half the cohort does not imply that the remainder lacked adequate control: many infections do not require an intervention. The relevant boundary is the exclusion of unresolved foci requiring longer treatment.^{1,2}

Reduced exposure is demonstrated; ecological benefit is not. Resistant organisms were detected using routine clinical samples rather than systematic surveillance, and *C. difficile* infections arising after discharge elsewhere could be missed. A patient without a positive clinical culture was therefore not necessarily free of resistant colonisation, and the trial did not measure population-wide transmission. Event rates and confidence intervals leave room for important differences in either direction. The observed resistance proportion was numerically higher with shorter treatment, but imprecise and exploratory. It would be equally misleading to claim that short treatment increased resistance or that BALANCE proved its prevention. The justification for reducing unnecessary antibiotics combines retained clinical effectiveness with less exposure, while acknowledging that downstream benefits were not established here.¹

Body of Evidence

Yahav and colleagues randomised 604 patients with Gram-negative bacteraemia to seven or fourteen days after they had become afebrile and haemodynamically stable. Uncontrolled infection sources were excluded, and 68% had a urinary source. The composite of death, infectious complications, readmission or prolonged admission at 90 days occurred in 45.8% versus 48.3%, meeting a ten-point noninferiority margin. This established a basis for shorter treatment in clinically stable patients, but its composite result did not precisely establish mortality noninferiority in a critically ill population.³

The Swiss trial led by von Dach compared a C-reactive-protein-guided strategy with fixed seven- and fourteen-day courses. It randomised 504 patients, with 503 in the analysed cohort, and required uncomplicated Gram-negative infection and absence of fever

before allocation. Clinical failure by day 30 occurred in 6.6% with seven days and 5.5% with fourteen days; the comparison met a ten-point noninferiority margin. That margin was large relative to the observed event rate. The biomarker arm also met noninferiority, but variable durations and nonadherence complicated interpretation.⁴

Molina and colleagues studied 248 patients with Enterobacterales bloodstream infection and appropriate source control. Their primary endpoint was antibiotic exposure, for which seven days was superior to fourteen. Clinical outcomes were assessed separately using a ten-point noninferiority margin; most met that criterion, although fever relapse did not. This distinction matters when summarising the earlier literature: the study was not a large mortality noninferiority trial. It supported reduced exposure with generally reassuring clinical outcomes in a selected population and permitted antibiotics to be resumed when needed.⁵

BALANCE changes this evidence base through both scale and case mix. It contains more participants than these three trials combined and makes death itself the primary endpoint. Its ICU cohort also includes more non-urinary infections than the ward cohort, reducing dependence on extrapolation from uncomplicated urinary bacteraemia. The investigators' supplementary meta-analysis combines the four trials, with BALANCE contributing 87.1% of the statistical weight. That synthesis is consistent with the absence of a major mortality penalty, but is largely driven by BALANCE and is not an independent replication of it.^{1,2}

Evidence from other infection syndromes supplies context rather than permission to use one duration universally. In STOP-IT, 518 patients with complicated intra-abdominal infection and adequate source control were assigned approximately four days of antibiotics or a longer course guided by resolution of physiological abnormalities. Median durations were four and eight days, with the composite of surgical-site infection, recurrent intra-abdominal infection or death occurring in 21.8% versus 22.3%. The result supported shorter treatment after adequate source control, but did not directly answer the bloodstream-infection question tested in BALANCE.⁶

These studies share a clinically important principle: duration cannot be interpreted separately from the infection source, adequacy of active treatment and need for further intervention. BALANCE used susceptibility-confirmed adequate therapy and excluded syndromes with established prolonged-treatment requirements. It did not randomise antibiotic choice, oral versus intravenous treatment, or source-control procedures. Its duration result therefore cannot validate an inactive agent, an unsuitable route or an unresolved focus. Those treatment decisions remain necessary before applying the shorter strategy.^{1,2}

Nor does seven days represent a proven biological minimum. BALANCE compared two practical plans, and the shorter plan substantially reduced exposure while retaining mortality outcomes within a reassuring interval. It did not establish whether selected patients could receive five days, whether biomarkers could improve individual stopping decisions, or whether particular excluded infections could safely be treated for less than their customary duration. Those are distinct questions requiring their own evidence rather than extrapolation from a successful seven-versus-fourteen-day comparison.¹

For practice informed by evidence available at the end of 2024, the trial shifts the default towards seven days of adequate treatment for eligible bloodstream infections, including many patients initially treated in ICU. Ongoing clinical review remains integral: persistent deterioration should prompt reassessment of the diagnosis, organism, drug exposure, source control and possible new infection. The evidence supports reducing routine prolonged treatment while retaining a reasoned response to complications, rather than allowing the calendar alone to determine care.¹

Can most eligible patients with bloodstream infection receive seven rather than fourteen days of antibiotics?

Yes. A planned seven-day course of adequate therapy is supported for patients resembling those enrolled, with reassessment for complications and respect for the trial's exclusions.

References

1. The BALANCE Investigators, for the Canadian Critical Care Trials Group, the Association of Medical Microbiology and Infectious Disease Canada Clinical Research Network, the Australian and New Zealand Intensive Care Society Clinical Trials Group, and the Australasian Society for Infectious Diseases Clinical Research Network. Antibiotic Treatment for 7 versus 14 Days in Patients with Bloodstream Infections. *N Engl J Med* 2025;392(11):1065–1078. Published online 20 November 2024.
2. BALANCE Investigators. Supplementary appendix to Antibiotic Treatment for 7 versus 14 Days in Patients with Bloodstream Infections. Published supplementary methods, tables and figures. *N Engl J Med*. Published online 20 November 2024.
3. Yahav D, Franceschini E, Koppel F, Turjeman A, Babich T, Bitterman R, et al.; Bacteremia Duration Study Group. Seven Versus 14 Days of Antibiotic Therapy for Uncomplicated Gram-negative Bacteremia: A Noninferiority Randomized Controlled Trial. *Clin Infect Dis* 2019;69(7):1091–1098.

4. von Dach E, Albrich WC, Brunel AS, Prendki V, Cuvelier C, Flury D, et al. Effect of C-Reactive Protein–Guided Antibiotic Treatment Duration, 7-Day Treatment, or 14-Day Treatment on 30-Day Clinical Failure Rate in Patients With Uncomplicated Gram-Negative Bacteremia: A Randomized Clinical Trial. *JAMA* 2020;323(21):2160–2169.
5. Molina J, Montero-Mateos E, Praena-Segovia J, León-Jiménez E, Natera C, López-Cortés LE, et al.; SHORTEN trial team. Seven-versus 14-day course of antibiotics for the treatment of bloodstream infections by Enterobacterales: a randomized, controlled trial. *Clin Microbiol Infect* 2022;28(4):550–557.
6. Sawyer RG, Claridge JA, Nathens AB, Rotstein OD, Duane TM, Evans HL, et al.; STOP-IT Trial Investigators. Trial of Short-Course Antimicrobial Therapy for Intraabdominal Infection. *N Engl J Med* 2015;372(21):1996–2005.

ADAPT-Sepsis

Dark P, Hossain A, McAuley DF, Brealey D, Carlson G, Clayton JC, et al.; ADAPT-Sepsis Collaborators. Biomarker-Guided Antibiotic Duration for Hospitalized Patients With Suspected Sepsis: The ADAPT-Sepsis Randomized Clinical Trial. JAMA 2025;333(8):682–693. Published online 9 December 2024.

Introduction

Starting antibiotics and stopping them are different decisions. In a patient with suspected sepsis, the consequences of delaying effective treatment may be immediate, whereas the point at which further treatment becomes unnecessary is often uncertain. Improving physiology, microbiological results and source control all inform that judgement, but none supplies a universal stopping date. Continuing treatment until every inflammatory abnormality resolves risks allowing the host response to determine antibiotic exposure long after the original indication has changed. Conversely, stopping because a laboratory result has improved may be inappropriate when an infection still requires treatment.

Serial procalcitonin measurements offer a possible additional signal. Earlier trials had shown reductions in antibiotic exposure, although their treatment decisions were generally unblinded and survival findings were less consistent. The 2021 Surviving Sepsis Campaign made a weak recommendation for combining procalcitonin with clinical evaluation when source control was adequate and the optimal duration remained unclear. C-reactive protein was a familiar alternative, but its routine availability did not establish the effectiveness of a formal stopping algorithm.¹

ADAPT-Sepsis tested both approaches within UK critical care. Its distinctive feature was to conceal the allocated biomarker strategy and laboratory results while giving clinicians daily advice about antibiotic discontinuation. This separated knowledge of a biomarker concentration from the decision support generated by it. The trial also counted antibiotic treatment over 28 days, allowing an apparent saving during the initial illness to be offset by subsequent treatment. Its central question was therefore practical and demanding: could an additional daily stopping signal reduce overall exposure while keeping mortality within an acceptable safety boundary? First published in December 2024, the trial belongs in this book despite its 2025 print citation.²

Synopsis

ADAPT-Sepsis was an investigator-initiated, three-group randomised trial in 41 NHS intensive care units. Recruitment ran from January 2018 to June 2024. Adults were eligible within 24 hours of starting intravenous antibiotics for suspected sepsis if treatment was expected to continue for at least 72 hours. Important exclusions included an anticipated course exceeding 21 days, severe non-sepsis-related neutropenia, expected treatment with an interleukin-6 receptor inhibitor, and a decision to discontinue active treatment within 24 hours. Clinicians identified suspected sepsis; microbiological confirmation was not required. Central allocation balanced site, sepsis severity and recent surgery.^{2,3}

Of 16,109 screened patients, 2,761 were randomised. One allocation made in error was removed, leaving 918 patients assigned to procalcitonin guidance, 924 to C-reactive protein guidance and 918 to standard care. Mean age was 60 years and 60% were male. Approximately 49% had septic shock and 28% had undergone surgery during the previous 72 hours. Respiratory infection predominated, with substantial abdominal and urinary infection populations. About two-thirds of infections were community acquired; an organism was identified in fewer than half. The cohort therefore included both confirmed infection and the diagnostic uncertainty common in routine sepsis care.²

All groups underwent daily research blood sampling. Laboratory staff knew which test, if any, to perform; patients, relatives, clinical teams and research staff did not receive the allocated group or biomarker concentrations. An automated system generated written advice. The control pathway produced advice supporting standard care, with an imposed delay intended to preserve concealment. Antibiotic choice, initiation and discontinuation remained clinical decisions. Routine C-reactive protein measurements were permitted outside the intervention when considered necessary, but were not to drive a separate formal stopping algorithm; routine procalcitonin use was restricted.^{2,3}

For procalcitonin, a concentration below 0.25 µg/L generated strong support for stopping antibiotics. A fall exceeding 80% from baseline, or a concentration between 0.25 and 0.50 µg/L, supported stopping; otherwise the advice supported standard care. For C-reactive protein, a concentration below 25 mg/L generated strong stopping advice, while a fall exceeding 50% from baseline supported stopping. These were advisory rules, interpreted alongside daily clinical assessment and microbiology. Research sampling and advice ended when treatment of the initial episode stopped, or at discharge, death or withdrawal. A later antibiotic course did not restart the intervention, although it contributed to follow-up antibiotic exposure.^{2,3}

The primary effectiveness outcome was total antibiotic duration from randomisation to day 28. The primary safety outcome was all-cause mortality at day 28, assessed against an absolute non-inferiority margin of 5.4 percentage points. The reported sample-size assumptions included a one-day antibiotic reduction and 15% control mortality. Each intervention was compared with the same control group. The corrected Methods specified a one-sided significance threshold of 0.025 for non-inferiority, together with a 95% confidence interval whose upper boundary did not exceed the margin. Effectiveness comparisons were not adjusted for multiplicity. Adjusted models, analyses excluding major deviations and analyses addressing missing data supplemented the principal results.²

Antibiotic duration was available for 898, 892 and 905 patients in the procalcitonin, C-reactive protein and control groups. Mean duration was 9.8, 10.6 and 10.7 days, respectively. The control-minus-procalcitonin difference was 0.88 days (95% CI 0.19 to 1.58; $P=0.01$). The corresponding C-reactive protein difference was 0.09 days (95% CI -0.60 to 0.79; $P=0.79$). Adjustment produced similar estimates. Treatment of the initial episode was shorter with both biomarkers: means were 7.0 days with procalcitonin, 7.4 with C-reactive protein and 8.1 with standard care. Only the procalcitonin saving remained evident across the complete 28-day period.^{2,3}

Day-28 mortality was known for 879, 874 and 878 patients. There were 184 deaths with procalcitonin, 184 with C-reactive protein and 170 with standard care: 20.9%, 21.1% and 19.4%. The procalcitonin-minus-control difference was 1.57 percentage points (95% CI -2.18 to 5.32; $P=0.02$ for non-inferiority), meeting the specified criterion. The C-reactive protein difference was 1.69 points (95% CI -2.07 to 5.45; $P=0.03$), which did not. In the analysis excluding major violations, procalcitonin again met the margin, with an upper confidence boundary of 5.02 points; the corresponding C-reactive protein boundary was 5.90 points. Neither result demonstrated mortality superiority.²

At 90 days, mortality was 25.6%, 26.4% and 25.5% among patients with available follow-up. There was no clear advantage in intensive care stay, discharge timing, unscheduled escalation or readmission, new infection, or suspected clinically relevant antibiotic-related events. Relapse or recurrence requiring further treatment affected 11 patients with procalcitonin, five with C-reactive protein and five controls; estimates were imprecise. Nine separately reported serious adverse events were distributed equally between groups and judged unrelated to the intervention. Expected sepsis complications were captured as outcomes rather than necessarily being recorded again as serious adverse events.^{2,3}

Implementation data showed more frequent and earlier strong stopping advice with procalcitonin. Among patients who received such advice, the mean time to its first production was 3.5 days with procalcitonin and 5.1 with C-reactive protein. These conditional summaries do not describe all randomised patients. Overall, 364 patients stopped the intervention but continued follow-up, while 127 withdrew completely before day 28. Sensitivity analyses supported the modest procalcitonin effect on antibiotic duration. A Bayesian analysis using an uninformative prior gave an 85% probability that its mean saving exceeded half a day. The investigators concluded that procalcitonin guidance reduced antibiotic duration and satisfied their mortality safety criterion; C-reactive protein guidance did not establish the same combination.^{2,3}

Critique

The strongest feature is the attempt to test decision support while concealing the laboratory signal that generated it. A conventional open trial allows clinicians to see a falling biomarker, know the treatment assignment and respond to expectations about what the study should achieve. ADAPT-Sepsis concealed both allocation and numerical results, collected samples in controls, and standardised communication. This reduces several opportunities for differential management. It also means the intervention was a complete laboratory-to-clinician advice system, not simply the availability of another blood test. Reproducing its effect requires attention to that clinical process.^{2,3}

Concealment was nevertheless imperfect in principle. Repeated stopping messages could suggest that a patient was receiving active guidance, whereas control advice never explicitly supported stopping. Clinicians might also defer a decision while waiting for advice that would never change in a control patient. The investigators acknowledged this possibility. Its magnitude cannot be determined from the observed control duration or comparisons with other studies. Routine C-reactive protein testing in all groups additionally makes that comparison one of formal algorithmic advice added to an environment where the biomarker was already available. The null result does not show that clinicians obtained no information from C-reactive protein at all.²

The antibiotic outcome was sensibly extended beyond the first treatment course. A stopping strategy that merely shifts antibiotic use into a later episode has achieved less than its initial-course duration suggests. The C-reactive protein findings demonstrate this distinction: an initial reduction of approximately seven-tenths of a day did not become a reduction in overall 28-day exposure. However, the trial did not establish why. Recurrence, new infections, other indications and subsequent prescribing all require consideration; the absence of a net saving should not automatically be attributed to microbiological relapse.²

Mortality is the more consequential limitation. An absolute margin of 5.4 percentage points would permit a sizeable increase from the observed 19.4% control mortality. The procalcitonin confidence interval excluded the selected margin only narrowly: its upper boundary was 5.32 points. Thus the trial met its statistical criterion while remaining compatible with a clinically important mortality increase. It also remained compatible with benefit. Describing this as proof that procalcitonin is harmless would discard most of the uncertainty the interval communicates. The result supports non-inferiority according to the chosen margin, rather than establishing negligible risk in exchange for the antibiotic saving.²

The two biomarker mortality estimates reinforce that point. Their point estimates and confidence intervals were very similar, despite falling on opposite sides of the decision boundary. One was labelled non-inferior and the other inconclusive; this does not establish a meaningful difference in safety between the biomarkers. Nor does $P=0.02$ indicate a mortality benefit. It belongs to the test against the non-inferiority margin. At 90 days, similar observed mortality provides useful additional context, but the estimates still lacked precision and follow-up was less complete. A nonsignificant later comparison does not erase the uncertainty at 28 days.²

Missing outcomes deserve attention because the safety conclusion was so close to its boundary. More than 4% withdrew completely, and the mortality analysis did not contain every randomised patient. Similar withdrawal counts are reassuring but cannot establish similar outcomes among those missing. The analysis excluding major violations supported procalcitonin non-inferiority, yet exclusions create their own selection and do not recover missing deaths. Patients who stopped receiving advice but retained follow-up were appropriately distinguished from those who withdrew altogether. The reported analyses are informative; none makes the mortality finding immune to assumptions about incomplete follow-up.^{2,3}

Death also competes with the opportunity to accumulate antibiotic days. A low total can reflect successful early discontinuation or early death, with very different clinical meanings. The investigators examined a win ratio prioritising death before antibiotic duration and obtained results broadly consistent with the effectiveness finding. That is preferable to ignoring the problem, but it remains an additional analysis of the same patients. It does not convert an exposure endpoint into an independently established survival or recovery benefit. Similarly, antibiotic duration and total defined daily doses are different measures; fewer treatment days should not be described as a demonstrated reduction in every dimension of antibiotic consumption.²

The average saving was real but modest, with uncertainty ranging from approximately one-fifth to one-and-a-half days. Its service value depends on baseline prescribing, assay and sampling requirements, timely delivery of advice and whether clinicians act on it. Benefits in resistance, adverse drug effects, workload or costs are plausible reasons to seek shorter treatment, but were not established by the principal clinical results. Several safety events were uncommon, and relapse was numerically more frequent with procalcitonin, with wide uncertainty. Neither the small event counts nor the nine separately classified serious adverse events justify a broad claim of safety across all relevant complications.^{2,3}

Generalisability is strongest for comparable adults early in suspected sepsis within a functioning critical care and stewardship service. The substantial shock and surgical populations are helpful, but the exclusions matter: long-course infections, severe neutropenia and interleukin-6 receptor inhibition cannot simply inherit the overall conclusion. Only 17% of screened patients entered the study, frequently because antibiotics had already exceeded the eligibility window. Subgroup plots provide no sufficiently convincing basis for selecting a particular infection phenotype for treatment. The later correction, published on 17 December 2025, clarified the non-inferiority wording and replaced the first plot of supplementary eFigure 3. This chapter uses that corrected record as an integrity update, without treating it as new 2025 efficacy evidence.^{2,3,4}

Body of Evidence

PRORATA was an important early test of this approach. It analysed 621 patients and allowed procalcitonin to inform both starting and stopping antibiotics, with final decisions retained by clinicians. Antibiotic-free days to day 28 increased from 11.6 to 14.3, a difference of 2.7 days. Day-28 mortality was 21.2% versus 20.4%; by day 60 it was 30.0% versus 26.1%. The latter difference had a 90% confidence interval extending to a 9.7-point increase, within the trial's 10-point non-inferiority margin. The study supported reduced exposure but illustrated how meeting a generous safety margin can coexist with appreciable uncertainty about mortality. Its open design and combined initiation/discontinuation strategy also differ from ADAPT-Sepsis.⁵

The Dutch SAPS trial subsequently analysed 1,546 patients receiving procalcitonin guidance or standard care. Advice supported discontinuation after an 80% fall from the peak concentration or a value of 0.5 µg/L or lower. Median treatment duration was five versus seven days, and defined daily dose consumption was lower. Mortality at 28 days was 149/761 versus 196/785, approximately 20% versus 25%. This favourable survival finding strengthened interest in procalcitonin, but the trial was open label and its

primary endpoints concerned antibiotic consumption and duration. ADAPT-Sepsis provides a useful independent test with concealed allocation and advice; its mortality results do not reproduce the same direction of benefit.⁶

A 2018 individual-patient meta-analysis assembled 4,482 participants from 11 trials. Procalcitonin guidance reduced antibiotic duration by an adjusted 1.19 days and was associated with lower mortality, 21.1% versus 23.7% (adjusted odds ratio 0.89; 95% CI 0.80 to 0.99). Those results placed an approximately one-day reduction within an established range of effects. However, the contributing trials did not blind patients and caregivers, and the synthesis included the earlier major trials rather than offering independent replication of them. ADAPT-Sepsis therefore adds substantially to the direct randomised evidence while making the survival interpretation more cautious than a simple continuation of the earlier pooled estimate might suggest.⁷

By 2021, the Surviving Sepsis Campaign had retained clinical evaluation as an essential partner to procalcitonin when considering discontinuation. Its recommendation was weak and based on low-quality evidence, applied after adequate source control, and concerned situations in which treatment duration remained uncertain. ADAPT-Sepsis strengthens the evidence for a modest effect on exposure in that broad clinical task. It does not supply an indication to delay necessary initial antibiotics, mandate discontinuation in an unstable patient or substitute a biomarker result for management of an uncontrolled focus. Those applications would move beyond the question tested.^{1,2}

The 2024 BALANCE trial offers a complementary way to reduce treatment. In selected bloodstream infections, seven days was non-inferior to 14 days for 90-day mortality, with important exclusions for infections requiring prolonged therapy. That comparison tested an intended duration rather than a biomarker threshold. ADAPT-Sepsis enrolled a different, diagnostically broader population and compared laboratory-guided advice with prevailing clinical practice. The results cannot establish which approach is preferable where both might apply. A clear evidence-based duration may leave little additional uncertainty for a biomarker to resolve, while patients outside a well-defined short-course population may present a different decision.^{2,8}

Taken together, the evidence supports procalcitonin as an optional aid to shortening treatment, with its greatest practical relevance where a real stopping decision remains unresolved. Adoption should be judged against the local standard of stewardship and accompanied by assessment of actual antibiotic exposure and patient outcomes. The biological appeal of individualisation is insufficient on its own; a useful service must deliver an interpretable result in time to change care. ADAPT-Sepsis established a modest average saving within its tested system and met its selected mortality margin. It

left uncertainty about smaller but important harms, broader patient benefit, and the added value of testing when clinicians already have a defensible stopping date.

Should daily procalcitonin guide antibiotic discontinuation in adults with suspected sepsis?

It can support a modest reduction alongside clinical assessment, but ADAPT-Sepsis only narrowly met its mortality non-inferiority margin. Routine C-reactive protein guidance did not reduce total antibiotic duration.

References

1. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med* 2021;47(11):1181–1247.
2. Dark P, Hossain A, McAuley DF, Brealey D, Carlson G, Clayton JC, et al.; ADAPT-Sepsis Collaborators. Biomarker-Guided Antibiotic Duration for Hospitalized Patients With Suspected Sepsis: The ADAPT-Sepsis Randomized Clinical Trial. *JAMA* 2025;333(8):682–693. Published online 9 December 2024.
3. ADAPT-Sepsis Collaborators. Supplement 3 to the ADAPT-Sepsis randomised clinical trial. *JAMA* 2025;333(8):682–693. Published supplementary methods, tables and figures.
4. Clarification of Noninferiority Wording and Corrected eFigure. *JAMA* 2026;335(2):187. Published online 17 December 2025.
5. Bouadma L, Luyt CE, Tubach F, Cracco C, Alvarez A, Schwebel C, et al.; PRORATA trial group. Use of procalcitonin to reduce patients' exposure to antibiotics in intensive care units (PRORATA trial): a multicentre randomised controlled trial. *Lancet* 2010;375(9713):463–474.
6. de Jong E, van Oers JA, Beishuizen A, Vos P, Vermeijden WJ, Haas LE, et al. Efficacy and safety of procalcitonin guidance in reducing the duration of antibiotic treatment in critically ill patients: a randomised, controlled, open-label trial. *Lancet Infect Dis* 2016;16(7):819–827.

7. Wirz Y, Meier MA, Bouadma L, Luyt CE, Wolff M, Chastre J, et al. Effect of procalcitonin-guided antibiotic treatment on clinical outcomes in intensive care unit patients with infection and sepsis patients: a patient-level meta-analysis of randomized trials. *Crit Care* 2018;22(1):191.
8. The BALANCE Investigators, for the Canadian Critical Care Trials Group, the Association of Medical Microbiology and Infectious Disease Canada Clinical Research Network, the Australian and New Zealand Intensive Care Society Clinical Trials Group, and the Australasian Society for Infectious Diseases Clinical Research Network. Antibiotic Treatment for 7 versus 14 Days in Patients with Bloodstream Infections. *N Engl J Med* 2025;392(11):1065–1078. Published online 20 November 2024.

Possible Reporting Inconsistencies. Table 2 reports the C-reactive protein comparison for suspected clinically relevant antibiotic-related events as an absolute difference of -0.84 with a 95% confidence interval of 3.43 to 9.44. The interval does not contain its point estimate and conflicts with the accompanying interpretation. No corrected unadjusted interval is inferred here; the adjusted supplementary estimate is a different analysis. The mortality and antibiotic-duration results are reported separately.

ASTER

Ware LB, Files DC, Fowler A, Aboodi MS, Aggarwal NR, Brower RG, et al.; for the National Heart, Lung, and Blood Institute Prevention and Early Treatment of Acute Lung Injury Clinical Trials Network. Acetaminophen for Prevention and Treatment of Organ Dysfunction in Critically Ill Patients With Sepsis: The ASTER Randomized Clinical Trial. JAMA 2024;332(5):390–400.

Introduction

Paracetamol is usually prescribed in critical care for pain or fever. ASTER addressed a different possibility: that scheduled administration might reduce the organ injury accompanying sepsis. The rationale concerned circulating cell-free haemoglobin, which can participate in oxidative reactions outside the protective environment of the red cell. Paracetamol can inhibit haemoprotein-mediated oxidation, suggesting a potential effect on tissue injury beyond temperature control. A familiar drug could therefore have an unfamiliar therapeutic role, but familiarity alone cannot establish the benefit or safety of giving it systematically to patients with acute organ dysfunction.¹

The preceding evidence was suggestive rather than definitive. Laboratory work supported the mechanism, and small clinical studies had raised the possibility of renal protection. These findings required confirmation in a larger and more varied sepsis population. There was also an important question about patient selection. If circulating haemoglobin was central to the mechanism, patients with higher concentrations might benefit more. Demonstrating that relationship would provide a reason to measure the biomarker before treatment, rather than exposing every patient with sepsis to the intervention.¹

ASTER was a phase 2b trial intended to test both the overall clinical effect and this potential route to identifying responders. It used intravenous treatment, a blinded placebo comparison, and an outcome combining survival with freedom from respiratory, circulatory and renal support. The main result did not establish an improvement in that outcome. Nevertheless, several exploratory respiratory findings attracted attention. The task in interpreting ASTER is to preserve the value of those signals for research while judging them in the context of the neutral primary result, multiple comparisons and an unconfirmed biomarker-treatment interaction.¹

Synopsis

The PETAL Network conducted ASTER in 40 US academic hospitals between October 2021 and April 2023. Adults with known or suspected infection and planned antibiotic treatment were eligible if they also required vasopressors after at least one litre of intravenous fluid, or had respiratory failure requiring assisted ventilation or at least 6 L/min of supplemental oxygen. Enrolment was permitted within 36 hours of presentation to an acute hospital, with admission or planned admission to a participating intensive care unit. Fever was not required. Exclusions recorded in the screening report included substantial liver enzyme elevation, cirrhosis, alcohol use disorder, long-term dialysis and several forms of chronic respiratory support.¹

The study began as a platform comparing intravenous paracetamol and intravenous vitamin C with a common placebo. The vitamin C comparison was stopped after external trial evidence became available, and the paracetamol comparison subsequently continued. The report presents the paracetamol comparison only. Web-based randomisation allocated treatment, and paracetamol and placebo infusions were identical in appearance. The reported cohort comprised 447 patients: 227 assigned to paracetamol and 220 to placebo. Mean age was 64 years, approximately half were female, 76% were receiving vasopressors and 42% assisted ventilation. Pneumonia accounted for about 44% of infections. Approximately 44% had already received paracetamol before enrolment.¹

The active regimen was 1 g intravenously every six hours for up to five days, reduced to 15 mg/kg per dose when actual body weight was below 50 kg. Placebo was 100 mL of 5% dextrose on the same schedule. Treatment stopped earlier at intensive care or hospital discharge, withdrawal or death. A new aminotransferase concentration reaching ten times the upper limit of normal prompted permanent discontinuation. Additional paracetamol was prohibited during study treatment, although a small number of patients received it. Other antipyretic drug use was discouraged, while physical cooling remained available. Supportive care recommendations included established lung-protective ventilation and a conservative fluid approach after shock resolution in patients with ARDS.¹

The primary endpoint was days alive and free of all three forms of organ support—assisted ventilation, vasopressors and new kidney replacement therapy—to day 28. The report distinguished this calendar-day count, which included survivors and nonsurvivors, from secondary free-day outcomes assigning zero to anyone who died by day 28. Individual organ-support components, mortality, intensive care and hospital outcomes, renal measures, SOFA scores and incident ARDS were also examined. The trial sought a 2.5-day primary-outcome difference, assuming a standard deviation of 8.7 days,

with approximately 218 patients per group. The main comparison used a two-sample t test. Analyses were exploratory and no multiplicity adjustment was applied.¹

Primary outcome data were available for 223 patients receiving paracetamol and 218 receiving placebo. Mean days alive and free of organ support were 20.2 and 19.6, a difference of 0.6 days (95% CI -1.4 to 2.6; P=0.56). The individual calendar-day components did not show a clear difference. For secondary free-day outcomes that assigned zero after death, ventilator-free days were 20.9 versus 19.4, vasopressor-free days 21.4 versus 20.0, and kidney replacement-free days 22.4 versus 21.5; all confidence intervals crossed no difference. Intensive care-free and hospital-free days were also similar. These measures should not be substituted for one another merely because their names sound alike.¹

At 28 days, all-cause mortality was 39/223 (17.5%) with paracetamol and 49/218 (22.5%) with placebo, an absolute difference of -5.0 percentage points (95% CI -12.4 to 2.5; P=0.19). At 90 days, all-cause mortality was 58/223 (26.0%) versus 69/218 (31.7%), again without a statistically clear difference. New kidney replacement therapy was initiated in 22/223 versus 18/218 patients. Neither the reported major adverse kidney event composite nor the change in creatinine established renal benefit. SOFA change from baseline to day seven was similar, although earlier daily comparisons favoured paracetamol for total and coagulation scores on days two to four and respiratory scores on days one to four.¹

Among patients assessed for new ARDS, four of 183 receiving paracetamol and 16 of 188 receiving placebo developed the syndrome within seven days: 2.2% versus 8.5%, an absolute difference of -6.3 points (95% CI -10.8 to -1.8; P=0.01). This was an incident-disease analysis rather than a comparison in all 447 patients; some already had ARDS at baseline. Overall initiation of assisted ventilation was similar, affecting 21/125 versus 23/132 eligible patients. Changes in the measured inflammatory, endothelial and cell-injury biomarkers did not show a consistent treatment effect by day three.^{1,2}

Baseline cell-free haemoglobin was available in 430 patients. The median concentration was approximately 8.4 mg/dL, and 44% exceeded the proposed 10 mg/dL threshold. Among those with both biomarker and primary outcome data, the estimated paracetamol effect was 2.0 days in the higher group and -0.2 days in the lower group; the interaction P value was 0.29. Mortality interaction testing was also inconclusive. An exploratory analysis suggested less new assisted ventilation in the higher-haemoglobin group with paracetamol, with an interaction P value of 0.02. A post hoc continuous biomarker analysis generated another possible mortality signal, but did not validate a treatment threshold.^{1,2}

Patients received a mean of 12.2 active doses or 12.8 placebo doses, fewer than the maximum 20 largely because treatment ended at intensive care discharge. Liver tests and measured fluid balance did not reveal a clear difference. Fluid administration or vasopressor initiation or escalation within two hours of an infusion occurred in approximately 40% and 41% of patients. There were 36 versus 32 reported adverse events, with patients potentially contributing more than one event. The investigators concluded that the regimen had no detected excess in the monitored safety outcomes, but did not improve the primary organ-support outcome. They considered the exploratory respiratory findings and haemoglobin-related hypotheses appropriate subjects for further research.^{1,2}

Critique

ASTER tested a coherent biological hypothesis using an appropriately blinded comparison. The intravenous route avoided uncertainty about enteral absorption, while an identical placebo helped separate pharmacological effects from changes in clinical behaviour. Recruitment across 40 hospitals broadened the evidence beyond a single centre. The substantial vasopressor population and inclusion of patients with respiratory dysfunction made the trial relevant to established acute illness rather than uncomplicated infection. Most importantly, it evaluated a clinical outcome rather than relying solely on movement in a biochemical marker. Those strengths make the neutral primary result consequential.¹

Neutral does not mean equivalent. The estimated effect was small, but its confidence interval extended from a 1.4-day disadvantage to a 2.6-day advantage, encompassing the targeted 2.5-day improvement at its upper end. The observed standard deviation of 10.6 days exceeded the 8.7-day design assumption. Thus the trial did not precisely exclude every clinically worthwhile effect. Mortality estimates were similarly uncertain: the point estimates favoured paracetamol, but the intervals included harm. Selecting the favourable mortality point estimate while dismissing its uncertainty would be no more defensible than declaring the treatment conclusively ineffective because *P* exceeded 0.05.¹

The outcome definition requires particular care. A calendar-day count incorporating nonsurvivors is not interchangeable with an endpoint that assigns zero free days to every patient who dies during follow-up. These approaches place different weights on survival and temporary recovery. The paper presents both, and neither yielded clear evidence of benefit. Combining respiratory, circulatory and renal support also makes the result dependent on the clinical mix and on which support lasts longest: a day qualifies

only when all three are absent. An improvement in one system can therefore fail to change the composite when another remains the limiting factor.¹

The ARDS result is the strongest reason for continuing investigation, but not a sufficient basis for routine treatment. It arose from only 20 incident events, in a subset without established ARDS and with available assessment, among numerous secondary comparisons. The absolute difference looks substantial, yet no adjustment protected against the chance of a favourable result emerging from this wider search. A phase 2 study can legitimately generate hypotheses in this way. The appropriate next step is a focused confirmatory test, not promotion of a secondary endpoint to the status of the original primary question.¹

The apparent agreement between incident ARDS and early respiratory SOFA scores adds biological interest, but those observations are not independent replications. Both describe respiratory illness in the same patients and may be influenced by related measurements and treatment decisions. The total SOFA score incorporates the respiratory component, further increasing overlap. Moreover, the day-seven SOFA comparison was neutral, and the numbers contributing daily scores fell as patients left observation through recovery, transfer or death. Differences among those still measured cannot automatically be interpreted as sustained benefit across the entire randomised cohort. The measured biomarker panel also did not establish target engagement. Inhibiting haemoglobin-driven oxidation need not lower the circulating haemoglobin concentration itself; an unchanged concentration therefore cannot, on its own, confirm or refute the proposed protective mechanism.^{1,2}

The proposed biomarker enrichment strategy was not confirmed. The relevant test is whether the treatment effect differs between higher and lower haemoglobin concentrations, not whether one subgroup alone has a favourable point estimate. That primary-outcome interaction was nonsignificant and imprecise. The assisted-ventilation interaction concerned a secondary event among patients not already receiving that support, with small numbers and multiple testing. Its direction also differed between biomarker groups. This is a useful candidate hypothesis, but it does not establish that a concentration above 10 mg/dL identifies patients who should receive paracetamol.^{1,2}

The continuous biomarker model is still further from a clinical decision rule. It was post hoc, and estimates at the extremes depend on relatively few observations and on the chosen functional form. The supplementary curve does not provide a validated threshold or show that applying a threshold improves outcomes. In addition, the proportion above 10 mg/dL was substantially lower than expected from earlier work using different assays. A numerical cut-off cannot be transferred between measurement systems

without establishing comparability. The trial obtained baseline samples for later analysis, rather than demonstrating a workable rapid-testing pathway that selected patients prospectively.^{1,2}

Exposure also limits the claim. The tested strategy was treatment during the intensive care stay for a maximum of five days, not five completed days for every participant. The average of approximately 12 doses is therefore partly a consequence of the stopping rule, rather than evidence that clinicians simply failed to deliver the intervention. Continuing treatment after intensive care discharge might produce a different result, but it was not tested. Previous paracetamol exposure in 44% could also reduce the contrast with an entirely untreated population. This does not invalidate randomisation, but it defines the practical comparison as adding scheduled study treatment after the care already received.¹

Safety findings are reassuring within the selected population and observed exposure, with similar hepatic and haemodynamic measures. They cannot establish unrestricted safety in patients excluded because of liver disease, marked enzyme elevation or alcohol-related risk. Rare toxicity remains difficult to exclude in a phase 2 sample, and a nonsignificant adverse-event comparison is not a formal safety non-inferiority result. Some supplementary safety tables use 199 placebo patients rather than the main cohort's 220, so their denominators require attention. The absence of demonstrated benefit also matters: tolerability alone does not justify four scheduled intravenous administrations daily as an organ-protective treatment for sepsis.^{1,2}

Body of Evidence

Experimental studies established the biochemical basis for the ASTER hypothesis. Bouteaud and colleagues showed that paracetamol inhibited haemoprotein-catalysed lipid oxidation in laboratory systems and attenuated renal injury in a rat model of rhabdomyolysis. This work supported a mechanism involving reactive haemoprotein intermediates rather than simply antipyresis. Its clinical implication was a reason to test the drug, not evidence that it would protect organs across the heterogeneous causes of human sepsis. Translation required demonstration of both an appropriate target population and a meaningful patient benefit.³

The small ACROSS study provided an early clinical test in severe sepsis with detectable cell-free haemoglobin. Forty patients satisfied treatment and sampling requirements for its principal analysis, although only 31 had the day-three biomarker measurement. The primary endpoint, day-three F_2 -isoprostanes, was not significantly reduced. A day-two comparison and several creatinine analyses favoured paracetamol. The selective analysis

population, missing samples and baseline differences limited certainty. It would therefore be inaccurate to describe ACROSS as an unequivocally positive trial of its primary endpoint. ASTER appropriately sought a larger clinical test, but did not reproduce convincing renal protection.⁴

A separate phase 2 trial in 62 patients with severe or moderately severe falciparum malaria compared paracetamol with no paracetamol alongside antimalarial treatment. The median creatinine reduction at 72 hours was 23% versus 14% ($P=0.043$), with a larger difference among patients with greater haemolysis. This supported further investigation in a disease where circulating haemoglobin has a particularly plausible role. The trial was small and open label, and its renal endpoint and disease context differ from ASTER. It cannot establish an indication for paracetamol in a general sepsis population whose organ dysfunction may have many other determinants.⁵

HEAT addressed the more familiar question of treating fever. Seven hundred patients with a temperature of at least 38°C and known or suspected infection were assigned intravenous paracetamol or placebo. Median intensive care-free days to day 28 were 23 versus 22, with a Hodges–Lehmann difference of zero days (96.2% CI 0 to 1; $P=0.07$). Ninety-day mortality was 15.9% versus 16.6%, without a clear difference. These findings did not establish improved recovery from routinely treating infectious fever. However, HEAT required fever and did not select treatment according to cell-free haemoglobin, so it did not directly settle ASTER’s mechanistic hypothesis.⁶

The studies consequently answer related but distinct questions. Fever reduction, inhibition of haemoprotein-mediated injury, renal protection in malaria and prevention of new ARDS are not interchangeable indications. A favourable finding in one setting should not be used to fill an evidential gap in another. Earlier observational associations between paracetamol exposure and survival also require caution: prescribing depends on symptoms, contraindications and illness trajectory. Randomisation addresses that selection problem, which is why those associations cannot outweigh a subsequent controlled comparison. ASTER’s contribution is its blinded multicentre comparison in a broader population with early sepsis and organ dysfunction. It showed that the tested regimen did not clearly improve its combined organ-support endpoint, while identifying respiratory observations worth pursuing in a more focused design.¹

Future work would need to decide whether the intended intervention is treatment of all eligible sepsis patients or treatment of a biologically selected subgroup. Those choices require different evidence. A prevention trial would need a clear population at risk of new ARDS and an outcome hierarchy that makes the respiratory question central. A biomarker-guided strategy would additionally need a reproducible assay, a justified

threshold and confirmation of a treatment interaction. ASTER supplies candidates for those decisions, but does not complete their validation. For current interpretation, paracetamol remains a drug with established symptomatic uses and an unproven additional role as routine organ protection in sepsis.

Should intravenous paracetamol be given routinely to prevent organ dysfunction in sepsis?

No. ASTER did not establish benefit on its primary organ-support outcome; the exploratory respiratory findings and proposed haemoglobin-defined treatment response require confirmation.

References

1. Ware LB, Files DC, Fowler A, Aboodi MS, Aggarwal NR, Brower RG, et al.; for the National Heart, Lung, and Blood Institute Prevention and Early Treatment of Acute Lung Injury Clinical Trials Network. Acetaminophen for Prevention and Treatment of Organ Dysfunction in Critically Ill Patients With Sepsis: The ASTER Randomized Clinical Trial. *JAMA* 2024;332(5):390–400.
2. PETAL Network ASTER Investigators. Supplement 2 to the ASTER randomised clinical trial. *JAMA* 2024;332(5):390–400. Published supplementary tables and figures.
3. Boutaud O, Moore KP, Reeder BJ, Harry D, Howie AJ, Wang S, et al. Acetaminophen inhibits hemoprotein-catalyzed lipid peroxidation and attenuates rhabdomyolysis-induced renal failure. *Proc Natl Acad Sci U S A* 2010;107(6):2699–2704.
4. Janz DR, Bastarache JA, Rice TW, Bernard GR, Warren MA, Wickersham N, et al.; Acetaminophen for the Reduction of Oxidative Injury in Severe Sepsis Study Group. Randomized, placebo-controlled trial of acetaminophen for the reduction of oxidative injury in severe sepsis: the Acetaminophen for the Reduction of Oxidative Injury in Severe Sepsis trial. *Crit Care Med* 2015;43(3):534–541.
5. Plewes K, Kingston HWF, Ghose A, Wattanakul T, Hassan MMU, Haider MS, et al. Acetaminophen as a Renoprotective Adjunctive Treatment in Patients With Severe and Moderately Severe Falciparum Malaria: A Randomized, Controlled, Open-Label Trial. *Clin Infect Dis* 2018;67(7):991–999.

6. Young P, Saxena M, Bellomo R, Freebairn R, Hammond N, van Haren F, et al.; HEAT Investigators and the Australian and New Zealand Intensive Care Society Clinical Trials Group. Acetaminophen for Fever in Critically Ill Patients with Suspected Infection. *N Engl J Med* 2015;373(23):2215–2224.

Possible Reporting Inconsistencies. The published flow diagram contains inconsistent allocation and follow-up counts. This chapter uses the main text’s cohort of 447 patients and Table 2’s primary analysis of 441. The Methods describe the renal composite using alternative components, whereas Table 2’s footnote joins creatinine deterioration and dialysis with ‘and’. Several supplementary safety tables use 199 placebo patients rather than 220. These reporting differences limit reconstruction of the exact populations and renal definition; the reported primary outcome is presented without inferring corrections.

REGARD-VAP

Mo Y, Booraphun S, Li AY, Domthong P, Kayastha G, Lau YH, et al.; on behalf of the REGARD-VAP investigators. Individualised, short-course antibiotic treatment versus usual long-course treatment for ventilator-associated pneumonia (REGARD-VAP): a multicentre, individually randomised, open-label, non-inferiority trial. Lancet Respir Med 2024;12(5):399–408.

Introduction

Stopping antibiotics for ventilator-associated pneumonia is difficult because both the diagnosis and the evidence of resolution are imperfect. Respiratory cultures may identify colonising organisms, radiographic abnormalities can persist after infection improves, and fever or organ dysfunction may have another cause. Continuing treatment can therefore seem safer than stopping, particularly when resistant Gram-negative organisms have been isolated. However, additional antibiotic exposure also carries costs: drug toxicity, further opportunities for resistant organisms to emerge, and treatment that may continue without a clear remaining indication.

The duration question has two distinct forms. One asks whether a shorter fixed course can replace a longer fixed course. The other asks whether a patient's clinical response can determine when treatment ends. The second approach might spare early responders several potentially unnecessary days while retaining longer treatment for patients whose infection remains uncontrolled. It requires a reproducible stopping rule and evidence that applying that rule preserves clinically important outcomes. Apparent recovery alone does not establish that all short courses are adequate.

REGARD-VAP evaluated a practical response-guided strategy using temperature and haemodynamic stability. It included hospitals in Nepal, Singapore and Thailand, with substantial representation of resistant Gram-negative infection. Patients were randomised after meeting criteria for stopping antibiotics, making the trial directly relevant to a decision clinicians face after initial treatment has worked. Its principal finding supported non-inferiority against substantially longer usual treatment. Interpretation depends on preserving that population and comparator, understanding the accepted margin of uncertainty, and separating the primary clinical result from striking but exploratory toxicity findings.¹

Synopsis

REGARD-VAP was a multicentre, individually randomised trial conducted in 39 intensive care units within six hospitals between May 2018 and December 2022. Adults had received invasive ventilation for at least 48 hours and met clinical and radiographic criteria for ventilator-associated pneumonia. Positive respiratory cultures were not mandatory. Exclusions included a SOFA score above 11, substantial immunocompromise, and concurrent infections requiring more than seven days of antibiotics. The report separately describes excluding a Brazilian site's single randomised patient because recruitment was minimal and data quality monitoring had not occurred; the published cohort concerns the six Asian hospitals.¹

Treatment duration was counted from the start of antibiotics covering at least one recovered respiratory pathogen according to susceptibility testing. For culture-negative episodes, antibiotic selection followed local microbiology or prescribing guidance. Patients were reviewed daily and became eligible for randomisation after temperature remained below 38.3°C orally or rectally, or below 38.0°C axillary, for 48 hours, with systolic blood pressure at least 90 mm Hg without inotropic support. The intended intervention therefore followed clinical improvement rather than applying an automatic early stop to every newly diagnosed episode.¹

Allocation used computer-generated variable blocks, stratification by site and sequentially numbered opaque envelopes. In the short-course group, antibiotics could stop from day three for culture-negative infection or day five for culture-positive infection, and were to stop within seven days. Usual care required at least eight days, with the precise duration chosen by treating clinicians. Investigators confirmed clinicians' willingness to follow allocation and provided stopping reminders. Clinicians knew the assignment; patients were not told their treatment duration, and independent specialists assessing recurrent pneumonia were masked to allocation and antibiotic duration.¹

Among 10,979 screened intensive care admissions, 461 patients were randomised: 232 to short treatment and 229 to usual care. One withdrew permission to use their data, leaving 231 and 229 in the reported intention-to-treat population. All 460 completed follow-up. The per-protocol population comprised 211 and 224 patients, excluding 18 with treatment-duration non-adherence and seven who did not satisfy enrolment criteria. Median age was 64 years, 39% were female, 81% were recruited in Thailand and approximately 69% were in surgical intensive care units. Median SOFA score was six, and ventilation had preceded pneumonia symptoms by a median of 14 days.¹

Respiratory cultures were positive in 320 patients and negative in 140. There were 491 isolates, 94% Gram-negative; these isolate percentages should not be read as percentages of patients. Non-fermenting Gram-negative organisms were identified in 218 patients, and carbapenem-resistant Gram-negative organisms in 141. Almost all positive respiratory specimens were endotracheal aspirates. Combination treatment was used in 47% of patients, and colistin- or polymyxin-B-based combinations were common for carbapenem-resistant infection. The median index antibiotic course was six days with the short strategy and 14 days with usual care. Nevertheless, 72 short-course patients, 31%, restarted antibiotics within five days.^{1,2}

The primary endpoint combined death or recurrent pneumonia within 60 days. Recurrence required agreement from two independent specialist assessors. Follow-up was daily during antibiotic treatment, then weekly while patients remained in hospital. After discharge, further assessments were scheduled at days 28 and 60 from enrolment. Suspected recurrence based on only one assessor was examined separately. The trial used a hierarchical analysis: non-inferiority first, followed by superiority only if non-inferiority was established. The absolute non-inferiority margin was 12 percentage points. The investigators expected a 55% control event rate and planned enrolment of 460, allowing for loss to follow-up. Non-inferiority required the upper one-sided 95% confidence bound to remain below the margin in unadjusted and adjusted analyses of both intention-to-treat and per-protocol populations. Adjusted analyses used baseline covariates, with g-computation for intention-to-treat and inverse probability weighting for per-protocol comparisons.¹

The primary endpoint occurred in 95/231 patients, 41%, with short treatment and 100/229, 44%, with usual care. The reported unadjusted risk difference was -3 percentage points, with a one-sided 95% upper confidence bound of +5 points. In the per-protocol population, the corresponding counts were 87/211 and 99/224, again a difference of -3 points and upper bound of +5. Adjusted upper bounds were +5 and +4 points. All four analyses met non-inferiority; none established superiority. A supplementary analysis using two-sided 95% intervals gave upper bounds of +6 to +7 points, also below the 12-point margin.^{1,2}

Sixty-day mortality was 81/231, 35%, versus 88/229, 38%; pneumonia recurrence occurred in 33/231, 14%, versus 30/229, 13%. These component counts overlap because a patient could experience recurrence and subsequently die. Subgroup interaction tests did not establish a differential composite treatment effect for carbapenem-resistant organisms, non-fermenting Gram-negative organisms, or culture status. Among patients with non-fermenters, recurrence occurred in 20/113 with short treatment and 13/105 with usual care; among those with carbapenem-resistant organisms, the corresponding

counts were 13/76 and 6/65. These numerically higher recurrence counts require interpretation alongside uncertainty and the composite analysis.^{1,2}

In the per-protocol population, mean antibiotic exposure during hospitalisation was 20.5 versus 25.7 days, an adjusted difference of –5.2 days (95% CI –7.5 to –2.8). Mechanical ventilation duration, intensive care stay and hospital stay were similar. Post-hoc assessment identified antibiotic side-effects in 17/211 patients, 8%, versus 86/224, 38%, with an adjusted difference of –31 percentage points (95% CI –37 to –25). Acute kidney injury accounted for most of this difference: 11/211 versus 79/224. New carbapenem-resistant colonisation or infection occurred in 37/211 versus 41/224. The investigators concluded that their clinical-response strategy reduced exposure while meeting the primary non-inferiority criterion.¹

Critique

REGARD-VAP's major strength is its focus on an actionable decision supported by measurements available in most hospitals. It did not require a proprietary assay, bronchoscopy or a complex score to decide when treatment could end. Randomisation after patients met stopping criteria protected the comparison from clinicians treating the initial episode differently because of the future allocation. Complete follow-up in the retained cohort and masked recurrence assessment further strengthened the primary result. The large separation in actual antibiotic duration shows that the study changed treatment meaningfully, rather than comparing two groups that ultimately received similar courses.¹

The eligibility process defines the limits of that result. Patients had to survive and improve sufficiently to enter the randomised comparison. The screening report identifies 564 admissions that did not meet stopping criteria by day seven, alongside exclusions for severe organ dysfunction, immunocompromise and other infections. The trial therefore supports a decision among early responders, not cessation at day seven despite persistent shock or clinical deterioration. Selection before randomisation is appropriate for this question, but it precludes estimating the effect of applying the rule indiscriminately from the moment pneumonia is suspected.¹

The control regimen is equally important. Usual care required at least eight days and delivered a median of 14. Consequently, this was not a direct comparison of individualised stopping against a consistent seven-day policy. Much of the reduction may reflect avoiding prolonged treatment after improvement. A unit already treating comparable patients for seven days would have less exposure to remove, and the trial does not quantify its additional gain from stopping at three to six days. Likewise, the

observed median of six days is a description of a strategy, not a universally validated optimum.¹

The 12-percentage-point non-inferiority margin permitted a substantial increase in a serious composite outcome. Its clinical acceptability should be considered separately from whether the statistical test succeeded. Nevertheless, the actual upper bounds were appreciably smaller than 12 points: approximately five using the reported one-sided 95% intervals, and seven using two-sided 95% intervals. This distinction matters because a one-sided 95% bound is less stringent than the upper bound of a conventional two-sided 95% interval. Both analyses support the stated non-inferiority conclusion, while still allowing a clinically relevant excess of events. Neither establishes equivalence or a mortality advantage.^{1,2}

Combining death and recurrence has reasonable clinical motivation, but can conceal differences between its components. Death prevents a later recurrence from being observed, and the two events have different importance. Overall mortality numerically favoured short treatment, whereas recurrence was slightly more frequent. In the resistant and non-fermenter subgroups, recurrence was also numerically higher with the short strategy. Nonsignificant interaction tests address evidence of treatment-effect variation; they do not prove organism-specific non-inferiority or eliminate a recurrence concern. The subgroup samples and confidence intervals are insufficient for a blanket claim that every resistant pathogen can receive the same abbreviated course.^{1,2}

Clinical diagnosis makes the findings pragmatic but introduces heterogeneity. Thirty per cent of episodes were culture-negative, and positive endotracheal cultures cannot invariably distinguish colonisation from invasive infection. Some patients may have improved because the original diagnosis was incorrect or because an alternative process resolved. Randomisation should distribute this uncertainty between groups, preserving the policy comparison, but the result cannot define the minimum treatment needed for unequivocal pneumonia caused by each organism. Moreover, duration counted from coverage of at least one isolate does not demonstrate equally adequate activity against every organism in a polymicrobial specimen.¹

The 31% early antibiotic restart rate prevents interpreting the intervention as uninterrupted freedom from antibiotics after the first stop. Restarting may reflect another infection, suspected recurrence or clinician concern; the published summary does not justify assigning all restarts to treatment failure. The more informative stewardship endpoint includes subsequent exposure: total hospital antibiotic days remained lower. The findings support stopping with continued clinical reassessment and

the ability to restart when indicated, rather than a promise that the initial short course will be the patient's final antibiotic treatment.¹

Per-protocol agreement adds reassurance but does not reproduce the protection of randomisation. Exclusions were more frequent in the short group, and decisions to continue treatment can depend on prognosis. Weighting adjusted for measured baseline characteristics; it cannot guarantee removal of unmeasured or evolving reasons for non-adherence. The supplementary weighting diagnostics showed modest variability, but a stable set of weights does not establish that all confounding was captured. Requiring consistent conclusions across four related analyses is useful, although those analyses remain observations from the same trial.^{1,2}

The toxicity result deserves particular restraint. It was a post-hoc assessment, presented prominently in the per-protocol population, and the reported observation window ended with index antibiotic treatment. Patients receiving longer courses therefore had both longer exposure and a longer opportunity to meet an on-treatment toxicity definition. Creatinine changes are objective, but their attribution to antibiotics is less certain in critically ill patients. The very large acute kidney injury difference is compatible with benefit from avoiding nephrotoxic treatment, yet cannot be treated as an equally secure primary randomised estimate of kidney protection. Its size may also depend on local use of polymyxins and other older regimens.¹

Finally, multicountry participation broadens relevance, but recruitment was dominated by Thailand and by referral hospitals. Similar country-specific point estimates from small samples do not establish equal performance everywhere. No clear reduction in newly acquired resistant organisms was demonstrated, and unit-wide colonisation or transmission was not measured. Reduced prescribing is an established result; reduced ecological resistance remains a plausible downstream benefit requiring separate evidence. Implementation should preserve the trial's clinical improvement, microbiology review and follow-up safeguards while evaluating local antibiotic exposure and recurrence.¹

Body of Evidence

The PneumA trial provided an influential fixed-duration comparison in 401 patients with microbiologically confirmed ventilator-associated pneumonia and appropriate initial empirical treatment. Eight versus 15 days produced 28-day mortality of 18.8% versus 17.2%, with more antibiotic-free days after the shorter course. Among patients with non-fermenting Gram-negative infection, pulmonary recurrence was higher with eight days: 40.6% versus 25.4%, a difference of 15.2 points (90% CI 3.9 to 26.6). That subgroup finding explains continuing concern about abbreviated treatment for these organisms. REGARD-VAP added a response-based selection step and a different microbiological setting; it did not simply repeat PneumA.³

Micek and colleagues tested an antibiotic discontinuation policy in a single medical intensive care unit, with 290 patients completing the trial. Mean treatment duration for suspected ventilator-associated pneumonia was six rather than eight days. Secondary pneumonia, hospital mortality and intensive care length of stay did not differ significantly. This earlier study showed that a stopping policy could reduce prescribing, but its size and single-centre setting limited certainty about uncommon harms and broader implementation. REGARD-VAP extended the strategy question to a larger multicentre non-inferiority comparison with longer follow-up.⁴

The 2016 ATS/IDSA guideline recommended seven days for ventilator-associated pneumonia, while allowing duration to vary according to improvement in clinical, radiographic and laboratory findings. It also acknowledged uncertainty about whether biomarker-guided stopping adds benefit when usual courses are already seven days or shorter. This places REGARD-VAP in context: the principle of shorter treatment was established before 2024, while a practical clinical-response rule and evidence from settings with considerable resistance remained valuable additions. The longer control treatment makes the trial especially relevant where prolonged courses persist.⁵

iDIAPASON specifically examined *Pseudomonas aeruginosa* pneumonia. Recruitment stopped early, leaving 186 patients. In the corrected report, the composite of death or *Pseudomonas* recurrence during intensive care follow-up occurred in 35.2% after eight days and 25.5% after 15 days: difference 9.7 points (90% CI –1.9 to 21.2). Non-inferiority was not established, and recurrence was 17.0% versus 9.2%. Failure to establish non-inferiority is not proof that the shorter course is inferior, but these uncertain findings remain relevant. REGARD-VAP's broader non-fermenter subgroup, selected after clinical improvement, cannot fully settle the narrower *Pseudomonas* question.⁶

A REGARD-VAP economic analysis published in November 2024 used the same trial data within country-specific decision models. Estimated probabilities that short treatment

was cost-effective were approximately 50% in Nepal, 55% in Singapore and 61% in Thailand; uncertainty intervals for net monetary benefit crossed zero in all three settings. These results support considering economic value, while showing that the estimates were far from decisive. They are modelled consequences of the original clinical evidence, not an independent replication or proof of certain savings in every hospital.⁷

Together, the evidence supports avoiding prolonged antibiotic treatment by default once a suitable patient has improved. The remaining question is how confidently a response-based rule can shorten an already brief course, particularly for specific pathogens or patients outside the studied eligibility criteria. A further comparison against a consistent seven-day strategy would address that gap more directly. Such research should count subsequent antibiotic courses, adjudicate recurrence independently and assess kidney outcomes over an equal follow-up period. These choices would help distinguish benefits of reduced drug exposure from differences in outcome observation, while preserving the practical decision that clinicians need to make. For REGARD-VAP itself, the strongest conclusion is that a monitored, individualised stopping strategy substantially reduced antibiotic exposure while satisfying its composite non-inferiority test in selected adults. The recurrence, renal and economic findings refine that conclusion without turning it into an unrestricted rule for all ventilated patients.¹

Can antibiotics be stopped early after clinical improvement in ventilator-associated pneumonia?

Yes, in patients resembling those enrolled in REGARD-VAP, a monitored response-guided course of seven days or less was non-inferior to longer treatment; persistent instability and other indications for antibiotics require separate assessment.

References

1. Mo Y, Booraphun S, Li AY, Domthong P, Kayastha G, Lau YH, et al.; on behalf of the REGARD-VAP investigators. Individualised, short-course antibiotic treatment versus usual long-course treatment for ventilator-associated pneumonia (REGARD-VAP): a multicentre, individually randomised, open-label, non-inferiority trial. *Lancet Respir Med* 2024;12(5):399–408.
2. REGARD-VAP investigators. Supplementary appendix 3 to the REGARD-VAP trial. *Lancet Respir Med* 2024;12(5):399–408. Published supplementary results, pp1–19.

3. Chastre J, Wolff M, Fagon JY, Chevret S, Thomas F, Wermert D, et al.; for the PneumA Trial Group. Comparison of 8 vs 15 days of antibiotic therapy for ventilator-associated pneumonia in adults: a randomized trial. *JAMA* 2003;290(19):2588–2598.
4. Micek ST, Ward S, Fraser VJ, Kollef MH. A randomized controlled trial of an antibiotic discontinuation policy for clinically suspected ventilator-associated pneumonia. *Chest* 2004;125(5):1791–1799.
5. Kalil AC, Metersky ML, Klompas M, Muscedere J, Sweeney DA, Palmer LB, et al. Management of adults with hospital-acquired and ventilator-associated pneumonia: 2016 clinical practice guidelines by the Infectious Diseases Society of America and the American Thoracic Society. *Clin Infect Dis* 2016;63(5):e61–e111.
6. Bouglé A, Tuffet S, Federici L, Leone M, Monsel A, Dessalle T, et al.; on behalf of the iDIAPASON Trial Investigators. Comparison of 8 versus 15 days of antibiotic therapy for *Pseudomonas aeruginosa* ventilator-associated pneumonia in adults: a randomized, controlled, open-label trial. *Intensive Care Med* 2022;48(7):841–849.
7. Cai Y, Booraphun S, Li AY, Kayastha G, Tambyah PA, Cooper BS, et al. Cost-effectiveness of a short-course antibiotic treatment strategy for the treatment of ventilator-associated pneumonia: an economic analysis of the REGARD-VAP trial. *Lancet Glob Health* 2024;12(12):e2059–e2067.

Possible Reporting Inconsistencies. Table 3 labels several small decimal risk differences as percentages, and its unadjusted liver-injury estimate does not reconcile with one event among 211 patients versus two among 224. No liver-protection effect is inferred here. The stopping criteria require no inotropic support, but Table 1 records such support in 5% of each group on the randomisation day; the relationship between this calendar-day measurement and the timing of eligibility assessment is unclear.

Landi-SEP

Rehberg S, Frank S, Černý V, Cihlár R, Borgstedt R, Biancofiore G, et al.; on behalf of the LANDI-SEP Study Group. Landiolol for heart rate control in patients with septic shock and persistent tachycardia. A multicenter randomized clinical trial (Landi-SEP). Intensive Care Med 2024;50(10):1622–1634.

Introduction

Tachycardia in septic shock can represent several different problems. A patient may need a faster heart rate to maintain cardiac output when stroke volume is inadequate. Alternatively, persistent adrenergic stimulation, fever, pain or an arrhythmia may drive the rate beyond what the circulation requires. These mechanisms can coexist and change during treatment. A numerical heart rate threshold therefore identifies a physiological observation without necessarily identifying a therapeutic target. The challenge is to distinguish patients whose circulation might benefit from slowing from those who depend on their tachycardia.¹

Beta blockade offers a plausible way to interrupt excessive adrenergic stimulation, lengthen diastolic filling and reduce myocardial oxygen demand. Its opposing effects are equally relevant: reducing heart rate and contractility may compromise cardiac output in a patient whose blood pressure already requires substantial vasopressor support. A short duration of action allows rapid adjustment, but does not remove that physiological trade-off. A well maintained arterial pressure can also conceal an increased requirement for vasoconstriction or a reduction in flow. Successful rate control and successful shock treatment are consequently distinct claims.

Landiolol is an ultra-short-acting, highly selective β_1 antagonist. Landi-SEP investigated whether it could achieve sustained rate control in patients with persistent tachycardia after initial treatment of septic shock, while avoiding an increase in vasopressor dose at a specified assessment point. The trial deliberately combined a heart rate target with a measure intended to guard against haemodynamic deterioration. Published online on 19 September 2024, it provides useful evidence about the feasibility of this strategy. Its clinical interpretation depends on separating the components of that endpoint, examining the precision of patient outcomes and establishing how confidently clinicians can recognise the proposed responsive phenotype.^{1,2}

Synopsis

Landi-SEP was an open-label, randomised phase IV trial conducted at 20 recruiting sites in seven European countries between February 2018 and February 2022. Adults with septic shock remained eligible after 12–36 hours of haemodynamic optimisation if they still required noradrenaline and had a persistent heart rate of at least 95 beats/min. Investigators excluded patients whose tachycardia they considered compensatory. Other exclusions included a mean arterial pressure below 65 mmHg, cardiogenic shock, severe pre-existing cardiac failure, important untreated conduction abnormalities and circumstances in which beta blockade was contraindicated or anticipated independently of allocation. Previous chronic beta blocker treatment was not itself an absolute exclusion.^{1,2}

Of 2,079 patients screened, 200 were randomised: 99 to landiolol and 101 to usual care. A screening failure in the intervention group, two beta blocker related exclusions in the control group and one control withdrawal left 196 patients, 98 per group, in the principal analysis. Randomisation used a statistician-generated list and stratification for atrial fibrillation during the optimisation period. The primary comparison accounted for atrial fibrillation and centre. Although recruitment was multinational, five sites contributed 150 patients and Czech sites contributed 106 of the 200 randomised participants.^{1,2}

Mean age was approximately 64–65 years. Baseline mean SOFA scores were 12.6 and 12.1, with noradrenaline-equivalent doses around 0.51 and 0.52 µg/kg/min. Heart rates averaged 116.0 and 114.2 beats/min, mean arterial pressures approximately 79 mmHg and lactate concentrations around 4 mmol/L. Mechanical ventilation was used in 85% and 80%, and kidney replacement therapy in 24% and 22%. Respiratory and abdominal infections predominated. Fifty patients were in the atrial fibrillation randomisation stratum, 26 assigned to landiolol and 24 to control. This was a substantially supported shock population despite the requirement for initial optimisation.^{1,2}

Landiolol was started within two hours of randomisation at 1 µg/kg/min. Increments of 1 µg/kg/min could be made at intervals of at least 20 minutes, to a maximum of 40 µg/kg/min, aiming for 80–94 beats/min. Treatment could continue until vasopressor cessation, a relevant serious adverse event, death, critical care discharge or day 28. Actual exposure was much shorter: median infusion duration was approximately 41 hours, and the median average infusion rate over the first 24 hours was about 4 µg/kg/min. Thus, neither the maximum permitted dose nor the maximum treatment period describes the usual exposure. Controls received standard sepsis management, including other antiarrhythmic treatment when clinicians considered it necessary.^{1,2}

The primary endpoint required three conditions within the first 24 hours. Patients had to reach the target heart rate for three consecutive hourly observations, allowing a rate below 80 when judged clinically unimportant. Control then had to be maintained, without three successive out-of-range observations under the specified rules. Finally, the noradrenaline-equivalent dose at 24 hours had to be no greater than at treatment initiation. This last condition compared two time points; it did not require vasopressor dose to remain unchanged throughout the intervening period. Clinical judgements about the relevance of bradycardia formed part of endpoint assessment.¹

The combined endpoint was achieved by 39 of 98 patients receiving landiolol, compared with 23 of 98 controls, or 39.8% versus 23.5%. The stratified absolute difference was 16.5 percentage points (95% CI 3.4–28.8; $P=0.013$). Heart rate attainment alone occurred in 74 versus 42 patients, and maintained heart rate control in 57 versus 29. By contrast, the vasopressor condition alone was met by 56 versus 65 patients, or 57.1% versus 66.3%; its difference was –9.2 percentage points (95% CI –22.0 to 4.4; $P=0.19$). The positive composite therefore accompanied substantially better rate control, without improvement in its vasopressor component.¹

Mean heart rate at 24 hours was 90.2 beats/min with landiolol and 101.0 with control. Separation remained evident on day two and subsequently diminished. Mean duration of vasopressor treatment was 5.64 versus 5.08 days. Cumulative noradrenaline exposure was numerically higher with landiolol, although distributions were highly skewed and this descriptive comparison does not establish a treatment effect. Mean blood pressure trajectories were broadly similar. These observations show that effective heart rate reduction did not translate into demonstrated vasopressor sparing.^{1,2}

At day 28, 43 of 98 landiolol patients and 39 of 97 controls had died, or 43.9% versus 40.2%. The reported absolute difference was 3.8 percentage points (95% CI –9.9 to 17.3; $P=0.60$). The time-to-event analysis likewise remained inconclusive, with a hazard ratio of 1.35 (95% CI 0.85–2.13). Among survivors at day 28, median critical care stay was 14.0 versus 13.9 days. No convincing improvement in organ dysfunction or lactate trajectories emerged. The number contributing SOFA measurements fell substantially over time, limiting later comparisons.^{1,2}

Adverse events affected 66 landiolol patients and 65 controls; serious adverse events affected 54 and 52. The results supplement recorded hypotension in 11 versus three patients, including five landiolol patients whose hypotension was considered treatment related. Four landiolol patients experienced five bradycardic episodes; three episodes in two patients were considered related to treatment. Two patients had serious events judged related to landiolol: one low cardiac output syndrome and one hypotension

event. Both resolved after treatment adjustment or cessation. Similar overall adverse-event totals therefore coexisted with clinically relevant events consistent with beta blockade.^{1,2}

Exploratory analyses considered rhythm and cardiac function. In the atrial fibrillation stratum, the combined endpoint occurred in nine of 26 versus four of 24 patients. Mortality was nine of 26 versus ten of 24, an absolute difference of -7.1 percentage points with a wide 95% CI from -32.0 to 18.7. These small comparisons did not establish a patient-outcome benefit in atrial fibrillation. Analyses by ejection fraction were smaller still and should be considered exploratory.^{1,2}

Critique

The principal strength of Landi-SEP is a clear demonstration of pharmacological efficacy under demanding clinical conditions. Patients had persistent tachycardia, substantial vasopressor exposure and frequent concurrent organ support. Landiolol produced prompt rate reduction with a titratable infusion at doses generally well below the permitted maximum. Multicentre randomisation supports a causal interpretation of the improvement in rate control. The trial also avoided assuming that a lower heart rate was automatically desirable: its endpoint explicitly incorporated a vasopressor condition. Reporting the separate components makes the limitations of that safeguard visible.¹

However, the endpoint answers a narrower question than whether landiolol improves shock. A patient could require more vasopressor for part of the first day, return to the starting dose by hour 24 and satisfy the vasopressor condition. Another could maintain blood pressure through greater vascular resistance despite reduced flow. Neither scenario is excluded by the composite. The proportion meeting the vasopressor component was numerically lower with landiolol, and the study did not demonstrate a reduction in vasopressor duration. It is therefore inappropriate to translate the primary result into a claim of improved haemodynamic stability or reduced catecholamine requirements.^{1,2}

Endpoint ascertainment also contained a subjective element. Investigators could accept heart rates below 80 beats/min when they considered them clinically unimportant, and treatment was unblinded. Heart rate and infused dose are objective measurements, but the interpretation of a low rate and decisions to adjust vasoactive support are not independent of allocation. A continuously measured rate does not make the entire composite objective. The design is persuasive that clinicians can achieve the specified target more often with landiolol; it is less persuasive that the resulting state necessarily represents physiological improvement.

The selection of patients without compensatory tachycardia is clinically sensible but difficult to reproduce. Investigators made that distinction after an optimisation period, without a validated rule that can be applied confidently across different shock phenotypes. The report does not provide the serial cardiac output information needed to determine whether rate reduction improved stroke volume sufficiently to preserve flow in individual patients. Excluding recognised compensation cannot guarantee that all remaining tachycardia is dispensable. This problem limits both mechanistic interpretation and translation of the eligibility criterion into routine bedside practice.^{1,2}

Patient-centred outcomes remain uncertain. The mortality confidence interval accommodates an absolute reduction approaching ten percentage points and an increase exceeding seventeen. A non-significant P value does not establish equivalent survival or exclude important harm. Conversely, the numerically higher death rate does not prove that landiolol caused excess mortality. The study was designed around a physiological endpoint and was not powered to settle survival. Time-to-event and fixed-time estimates describe different quantities; their differing numerical impressions should not be treated as a reporting contradiction or selectively used to imply reassurance.¹

Other secondary outcomes need similarly careful interpretation. Critical care length of stay was reported among day-28 survivors, conditioning the comparison on an event occurring after randomisation. It cannot establish an overall reduction in resource use. SOFA data were available for 68 and 71 patients on day two, falling to 22 and 30 by day seven and seven and two by day 28. Later averages therefore compare changing, highly selected subsets. A nominally higher day-two SOFA score with landiolol is not convincing evidence of organ harm among multiple comparisons, but neither do these incomplete trajectories demonstrate organ protection.²

The analysis population and treatment overlap introduce further qualifications. Four randomised participants were excluded before the main analysis, so the 196-patient population was not a complete intention-to-treat analysis of all 200 allocations. Usual care also contained active rate management: 35 controls received antiarrhythmic medication during the first 24 hours, compared with 16 landiolol patients. Fifteen controls were recorded as not reaching the day-28 study visit because of beta blocker treatment. This is not equivalent to losing all their survival information, but it shows that clinical decisions affected ongoing study participation. The trial estimates adding a landiolol strategy to adaptable usual care, rather than a pure comparison with no rate-control treatment.^{1,2}

Safety reporting warrants attention to denominators and attribution. The 166 versus 159 adverse events were event counts, not numbers of affected patients. Likewise, five treatment-related hypotension cases in the intervention group should not replace the eleven patients with hypotension of any attribution. Investigators' judgements that events were unrelated to the drug cannot establish haemodynamic safety, especially in an unblinded shock trial where deterioration has many plausible causes. Rapid reversibility is useful when an adverse response is recognised, but two serious related cardiovascular events show that short drug action does not eliminate clinically important risk.²

The subgroup findings do not identify a population ready for preferential treatment. Atrial fibrillation was a randomisation stratum, but that fact does not validate a survival interaction. Its fifty patients generated a mortality interval compatible with large effects in either direction. Opposite numerical directions in patients with and without atrial fibrillation are insufficient evidence of different treatment responses without a reliable interaction analysis. The high-ejection-fraction subgroup contained only 26 patients. Selection of either rhythm or ventricular phenotype for practice would require prospective confirmation with suitable clinical outcomes.^{1,2}

Reporting discrepancies further restrict subgroup interpretation. The supplementary forest plot labelled as the primary endpoint reproduces counts for heart rate attainment, rather than the three-part composite. It therefore cannot establish consistency of the actual primary treatment effect across subgroups. A footnote and a pooled subgroup total contain additional discrepancies, recorded below. These issues do not overturn the internally coherent overall primary result, but they argue against confident extrapolation from the supplementary analyses. Commercial funding and several sponsor-employed authors were disclosed; this makes independent replication particularly useful, without itself invalidating the randomised comparison.^{1,2}

Body of Evidence

Morelli and colleagues' 2013 single-centre esmolol trial established much of the interest in beta blockade during septic shock. Among 154 patients, treatment achieved the target rate and was associated with 28-day mortality of 49.4%, compared with 80.5% in controls. The adjusted mortality hazard ratio was 0.39 (95% CI 0.26–0.59). Survival was a secondary outcome in a phase II study, and the exceptionally high control mortality made the result difficult to generalise. Extensive haemodynamic monitoring and frequent levosimendan use also characterised the treatment environment. This was a compelling signal for further investigation, rather than a sufficient basis for routine treatment.³

J-Land 3S subsequently evaluated landiolol in Japanese patients with sepsis-related tachyarrhythmia. Of 151 randomised patients, 150 entered the primary analysis. At 24 hours, 55% receiving landiolol and 33% receiving conventional therapy had a heart rate of 60–94 beats/min; the adjusted absolute difference was 23.1 percentage points (95% CI 7.1–37.5). Its entry criteria included sinus tachycardia, atrial fibrillation and atrial flutter, with a different rate target and less prolonged required catecholamine exposure than Landi-SEP. Serious treatment-related events included hypotension and cardiovascular events. The study strengthened the evidence for rate reduction, without establishing that patients with sustained septic shock benefit from a universal target below 95 beats/min.⁴

Cocchi and colleagues' two-centre phase II esmolol trial provided a less encouraging haemodynamic result. Forty of 42 randomised patients received the study interventions. Mean noradrenaline-equivalent dose at six hours, the primary endpoint, was 0.30 µg/kg/min with esmolol and 0.21 µg/kg/min with usual care ($P=0.15$). Shock-free days also did not improve. Some biomarker and oxygen-consumption findings suggested biological effects, but these did not establish faster shock reversal. The study illustrates why plausible reductions in adrenergic activity cannot substitute for evidence of improved circulatory support requirements.⁵

STRESS-L was particularly relevant to established septic shock. It enrolled patients with tachycardia receiving noradrenaline for more than 24 hours, but stopped after 126 of a planned 340 participants because of futility and concern about possible harm. Mean SOFA score over 14 days was 8.8 with landiolol and 8.1 with standard care, a difference of 0.75 (95% CI –0.49 to 2.0). Day-28 mortality was 37.1% versus 25.4%, with an imprecise absolute difference of 11.7 percentage points (95% CI –4.4 to 27.8). Early stopping limits certainty about magnitude, but this randomised experience provides no support for routine treatment of persistent tachycardia in established shock.⁶

These trials address related but distinct interventions. Esmolol and landiolol cannot be assumed to have identical effects, and eligibility, target rates, monitoring, concurrent treatment and endpoint timing differ. Nevertheless, the sequence of evidence exposes a consistent interpretive problem: a drug can lower heart rate without reducing vasopressor exposure, organ failure or death. Landi-SEP adds a positive rate-based composite to this literature; it does not resolve the patient-outcome uncertainty or neutralise the concern raised by STRESS-L. Differences between control-group mortality rates across trials cannot determine which within-trial comparison is trustworthy.

By the end of 2024, a defensible conclusion was that landiolol can facilitate closely monitored rate control in selected patients, while routine suppression of persistent sinus

tachycardia as a sepsis treatment remained unsupported. That conclusion leaves open a separate clinical question about managing an arrhythmia that itself threatens haemodynamic function. Landi-SEP was not a definitive comparison of landiolol with alternative drugs for such arrhythmias. Future trials would need reproducible physiological selection, explicit assessment of cardiac flow and adequately powered patient outcomes. A responsive subgroup should be demonstrated prospectively, rather than inferred from a small favourable mortality split after a neutral overall clinical result.¹

Should landiolol routinely be used to lower heart rate in septic shock?

No. Landi-SEP improved rate control but did not establish reduced vasopressor requirements, organ failure or mortality; its exploratory subgroup findings are insufficient to guide routine selection.

References

1. Rehberg S, Frank S, Černý V, Cihlář R, Borgstedt R, Biancofiore G, et al.; on behalf of the LANDI-SEP Study Group. Landiolol for heart rate control in patients with septic shock and persistent tachycardia. A multicenter randomized clinical trial (Landi-SEP). *Intensive Care Med* 2024;50(10):1622–1634.
2. LANDI-SEP Study Group. Electronic supplementary material to Landi-SEP. *Intensive Care Med* 2024;50(10):1622–1634. Published supplementary methods and results, Tables 1–14 and Figure 1.
3. Morelli A, Ertmer C, Westphal M, Rehberg S, Kampmeier T, Ligges S, et al. Effect of heart rate control with esmolol on hemodynamic and clinical outcomes in patients with septic shock: a randomized clinical trial. *JAMA* 2013;310(16):1683–1691.
4. Kakehana Y, Nishida O, Taniguchi T, Okajima M, Morimatsu H, Ogura H, et al.; J-Land 3S Study Group. Efficacy and safety of landiolol, an ultra-short-acting β_1 -selective antagonist, for treatment of sepsis-related tachyarrhythmia (J-Land 3S): a multicentre, open-label, randomised controlled trial. *Lancet Respir Med* 2020;8(9):863–872.
5. Cocchi MN, Dargin J, Chase M, Patel PV, Grossestreuer A, Balaji L, et al. Esmolol to treat the hemodynamic effects of septic shock: a randomized controlled trial. *Shock* 2022;57(4):508–517.

6. Whitehouse T, Hossain A, Perkins GD, Gordon AC, Bion J, Young D, et al.; for the STRESS-L Collaborators. Landiolol and organ failure in patients with septic shock: the STRESS-L randomized clinical trial. *JAMA* 2023;330(17):1641–1652.

Possible Reporting Inconsistencies. Supplementary Figure 1 is labelled as the primary endpoint but shows 74/98 versus 42/98, the counts for heart rate attainment; the actual composite counts are 39/98 versus 23/98. Table 2's maintenance footnote repeats the full composite definition despite different row counts. Supplementary Table 13 gives seven and three primary events in the two high-ejection-fraction groups but an overall total of twelve. The chapter uses the main endpoint definition and group-specific counts, and does not infer subgroup consistency from the mislabelled figure.

SCREEN

Arabi YM, Alsaawi A, Alzahrani M, Al Khathaami AM, AlHazme RH, Al Mutrafy A, et al.; for the SCREEN Trial Group and the Saudi Critical Care Trials Group. Electronic sepsis screening among patients admitted to hospital wards: a stepped-wedge cluster randomized trial. JAMA 2025;333(9):763–773.

Introduction

Recognising deterioration on a hospital ward requires more than collecting observations. Abnormal physiology must be noticed, interpreted and communicated to someone able to act. Electronic records offer a way to automate part of that sequence, repeatedly assessing recorded information and prompting clinical review. Sepsis is an attractive target because delayed recognition may postpone investigation, antimicrobial treatment and organ support. Yet an electronic warning has no therapeutic effect by itself. Its value depends on the accuracy and timeliness of its inputs, how it reaches staff and the response it produces.^{1,2}

The quick Sequential Organ Failure Assessment score, or qSOFA, combines low systolic pressure, tachypnoea and altered consciousness. These findings can identify patients at increased risk, but do not establish that infection is the cause. A useful deterioration alert might consequently improve care in patients with several different diagnoses. Equally, an alert labelled as sepsis could encourage unnecessary fluids, antibiotics or investigations. Evaluating such a system requires measurement of patient outcomes and unintended consequences, alongside whether clinicians acknowledge the warning.¹

SCREEN examined this problem at the level of ward practice. It combined a qSOFA-based electronic alert with staff training, a required communication process and feedback on responses. Rather than randomising individual patients to different interfaces on the same ward, investigators randomised the sequence in which wards adopted the programme. This allowed the intervention to become part of routine care while retaining a randomised comparison during its staged introduction. Published online on 10 December 2024, despite its 2025 print citation, the trial is an important contribution to the evaluation of digital clinical support. Its apparent mortality benefit must be interpreted through the complexities of that design, the COVID-19 pandemic and the distinction between improving general recognition and treating sepsis specifically.^{1,2}

Synopsis

SCREEN was a stepped-wedge cluster randomised trial in five Ministry of National Guard–Health Affairs hospitals in Saudi Arabia, all using the same electronic medical record. Forty-five wards were allocated by a concealed, computer-generated sequence to nine rollout groups of five wards. After a two-month baseline period, successive groups were scheduled to activate the programme at two-month intervals. Recruitment ran from October 2019 to July 2021, with follow-up through October 2021. Two periods were extended to three months during the first COVID-19 wave because admissions had fallen substantially.¹

Eligible wards provided medical, surgical, oncology or mixed care. Patient eligibility began at age 14 years, reflecting admission to these adult ward services. Patients without a commitment to full life support on arrival were excluded; subsequent changes in goals of care did not remove patients from the study. Two initially non-operational wards were excluded after randomisation when they became critical care units during the pandemic. Other wards contributed no data during intervals when they functioned as critical care units. In the 43 participating wards, 62,378 patients were assessed and 2,323 excluded because of their initial treatment limitations, leaving 60,055 in the analysis.^{1,3}

The alert operated silently in control periods and was revealed during intervention periods. It was triggered when at least two qSOFA components were met within a 12-hour window: systolic pressure no higher than 100 mmHg, respiratory rate at least 22 breaths/min or Glasgow Coma Scale score below 15. Blood pressure and respiratory rate were normally documented every four hours, and Glasgow Coma Scale every twelve hours. The system assessed newly entered observations using recent values. This was repeated electronic surveillance of intermittent clinical documentation, rather than continuous physiological monitoring.¹

Revealed alerts generated messages for nurses and physicians and an audible and visual warning on the charge nurse's handheld device. Nurses were prompted to assess the patient and contact the covering physician; physicians documented whether they considered sepsis present. Training addressed assessment, recording observations and early sepsis management. Ward dashboards and monthly reports provided feedback on acknowledgment rates and timing. Diagnostic and therapeutic decisions remained with the clinical team. All five hospitals already had rapid response teams, but the supplementary description recorded no existing early warning score systems.^{1,2,3}

Patients were analysed according to their first admission ward, although the actual alert followed the ward's current assignment. A control patient could therefore receive revealed alerts after transfer or when the ward adopted screening. There were 29,442

patients in screening periods and 30,613 in control periods. Median age was 59 years in both groups, and approximately half were male. Screening-period patients more often occupied medical or mixed wards. COVID-19 occurred in 11.8% versus 5.2%, an important imbalance associated with the calendar timing of implementation.¹

The primary outcome was death in hospital within 90 days. The principal generalised linear mixed model accounted for study period, clustering within wards, hospital and patient COVID-19 status, with a log link to estimate relative risk and jackknife estimation of standard errors. Three smaller hospitals were combined in the hospital covariate to achieve model convergence. The sample-size calculation sought 65,250 patients to detect a reduction in mortality from 3.13% to 2.46%, allowing for substantial intracluster correlation. Secondary outcomes and process measures were assessed with related models; false discovery rates were reported for multiple comparisons.^{1,3}

Alerts occurred in 4,299 screening patients and 5,394 controls, or 14.6% versus 17.6%. Among screening patients with an alert, 73.4% had nurse acknowledgment and 73.8% physician acknowledgment. Physicians classified 1,105 of the 3,173 acknowledged cases as sepsis, or 34.8%. This was their clinical assessment, rather than an independently adjudicated diagnostic reference. More than four-fifths of screening patients with alerts already had antibiotics ordered before the first alert, compared with nearly three-quarters of controls. The system frequently identified patients whose infection treatment was already under way.¹

In the twelve hours following the first alert, lactate was measured in 14.2% versus 10.3%, with an adjusted relative risk of 1.30 (95% CI 1.16–1.45). Intravenous fluids were ordered in 14.4% versus 6.8%, adjusted relative risk 2.17 (95% CI 1.92–2.46). There was no significant adjusted increase in blood, respiratory or urine cultures during this interval. New antibiotic orders were also similar, at 18.0% versus 18.2%, adjusted relative risk 0.97 (95% CI 0.89–1.06). Analyses extending across the remaining ward stay found more investigations and treatment orders with screening, but these were post hoc and involved a different observation period.^{1,3}

There were 937 deaths among 29,442 screening patients and 961 among 30,613 controls, crude risks of 3.2% and 3.1%. The unadjusted relative risk was 1.01 (95% CI 0.93–1.11). In the primary adjusted analysis, screening reduced mortality, with a relative risk of 0.85 (95% CI 0.77–0.93; $P < 0.001$). The investigators reported a corresponding number needed to screen of 206 (95% CI 140–425). Sensitivity analyses favoured screening after additional case-mix adjustment, exclusion of control patients admitted near ward crossover and treatment of calendar period as a fixed effect. A separate 14-day in-hospital mortality analysis also favoured screening.^{1,3}

Secondary findings were mixed. Adjusted vasopressor use was lower, with a relative risk of 0.86 (95% CI 0.78–0.94). Code blue activation was higher, adjusted odds ratio 1.24 (95% CI 1.02–1.50), as was incident kidney replacement therapy, 1.20 (95% CI 1.11–1.31). The latter occurred in 1,294 of 28,842 versus 1,088 of 30,168 eligible patients, excluding those receiving treatment before or within 48 hours of ward arrival. New multidrug-resistant organism flags were less frequent, but new *Clostridioides difficile* flags were more frequent. Adjusted comparisons did not establish differences in critical care admission, mechanical ventilation or rapid response activation.^{1,3}

The reported effect was larger in oncology wards; no significant interaction was found by age, documented infection or COVID-19 status. Among patients who actually generated an alert, mortality was 12.7% versus 11.2%, adjusted relative risk 1.04 (95% CI 0.93–1.16). Thus, the positive result applied to the ward population as a whole, while the analysis restricted to alerted patients did not show a mortality benefit.^{1,3}

Critique

SCREEN addresses a clinically important implementation question with a design stronger than a simple comparison before and after introducing software. Randomising rollout order creates contemporaneous comparisons during intermediate periods and lets wards contribute experience under both conditions. Evaluating all eligible patients captures possible changes in ward practice beyond the patients who trigger an alert. Routine electronic data allowed a large study with complete ascertainment of the stated primary outcome. The result therefore deserves attention as evidence about a deployed clinical programme, rather than merely a score's ability to predict risk.¹

The central statistical point is that the crude and adjusted results answer different questions. Screening was more common later in the trial, when COVID-19 and other changes affected mortality. Simply pooling patients across these periods discards the structure of randomisation and compares groups with different calendar exposure. Adjustment is consequently integral to the stepped-wedge analysis, not an optional attempt to rescue a neutral trial. The reported 15% relative reduction should be presented as the primary model-based estimate, together with the almost identical crude risks. It is not a fifteen-percentage-point absolute survival improvement.^{1,3}

That necessity does not make every modelling assumption secure. COVID-19 altered diagnoses, staffing, bed use, treatment and admission thresholds in ways that a binary patient variable may not fully capture. Some wards became critical care units, changing the contributing clusters and their patient populations. Only 43 wards in five hospitals generated the information, despite the very large patient count; the reported

intraclass correlation was 0.23. Treating all 60,055 observations as independent would exaggerate precision. Pooling three smaller hospitals and selecting a convergent covariance structure were practical analytical choices, but they reinforce the need to examine sensitivity to the model.^{1,3}

The reported sensitivity analyses are reassuring within those limits. Treating period as a fixed effect produced an adjusted relative risk of 0.73, although with a wider interval of 0.57–0.94. Additional case-mix adjustment gave an estimate of 0.80, and excluding controls admitted within 90 days of ward crossover gave 0.82. These analyses make a purely accidental dependence on one specification less likely. They are not independent replications, and a proposed small-sample correction did not converge. The appropriate interpretation is a result robust across several available analyses, with residual uncertainty about complex secular effects.³

The intervention itself cannot be reduced to an alert threshold. Staff received education, explicit communication expectations and repeated performance feedback, with organisational involvement in implementation. These elements may have improved care even when qSOFA remained below two. Conversely, the benefit might depend on the particular electronic record, nursing workflow and capacity for clinical response. Installing the same three-variable rule elsewhere would not reproduce the tested intervention automatically. The five hospitals' common management and electronic infrastructure strengthen implementation consistency while limiting the diversity of settings represented.^{1,2,3}

Its description as sepsis screening also requires qualification. The alert did not require suspected infection, and only about one-third of acknowledged cases were considered septic by the physician. This does not establish a false-positive rate, because clinical acknowledgment was incomplete and no independent diagnostic standard was applied. Nor does a prognostic association prove diagnostic accuracy: alerted patients had a much higher risk of death, but deterioration from other causes can produce the same observations. SCREEN tests whether a nonspecific physiological warning and response process can improve ward outcomes; it does not establish qSOFA as a reliable stand-alone sepsis diagnosis.¹

The alert-only analysis is especially vulnerable to selection. Allocation affected whether an alert occurred: fewer screening patients generated any alert, although those who did had more repeated alerts. Training might have changed documentation, earlier care might have prevented some thresholds being crossed, or the observed populations might have differed for other reasons. Selecting patients on this post-allocation event no longer preserves a clean randomised comparison. Its neutral mortality estimate

therefore does not refute the primary result, but neither does it prove that the benefit occurred exclusively in patients without alerts. Claims about that mechanism require additional evidence.^{1,3}

The process findings leave the causal pathway unresolved. More lactate measurement and fluid ordering show a response to alerts, but do not demonstrate appropriate resuscitation or correction of hypoperfusion. An order is also not the same as administered volume or timely delivery. New antibiotic orders during the first twelve hours did not increase, and most alerted patients were already receiving treatment. SCREEN therefore cannot be described as proving that faster antibiotics mediated the survival effect. Later increases in antibiotic ordering do not establish better drug selection, dosing or source control, none of which was comprehensively adjudicated.^{1,3}

Potential harms deserve the same attention as mortality. Higher code blue activation could reflect more active escalation, more arrests or changes in recording; the outcome does not distinguish these explanations. Increased kidney replacement therapy might reflect kidney injury, treatment decisions, survival long enough to receive support or other differences. Moreover, its denominator excluded dialysis occurring within 48 hours after ward arrival, a period in which screening could already influence care. The comparison is consequently more complicated than an analysis of renal injury in all patients. Neither these possibilities nor multiplicity adjustment justifies dismissing the signal.^{1,3}

The microbiological outcomes are also imperfect measures of harm. They came from infection-control flags for isolation precautions after 48 hours and did not consistently distinguish colonisation from infection. More *Clostridioides difficile* flags should not be relabelled as independently adjudicated symptomatic disease, and fewer resistant-organism flags do not demonstrate reduced transmission. The simultaneous findings do not establish that one antibiotic strategy was safer, particularly when appropriateness and spectrum were not evaluated. Statistically detectable changes in antibiotic-free days and length of stay were small and should not be presented as major resource savings.^{1,3}

Finally, death after hospital discharge was outside the primary outcome. Ninety-day in-hospital mortality is a meaningful endpoint, but it cannot establish overall survival to day 90 or recovery after discharge. The exclusion of patients with pre-existing limitations on life support also narrows applicability. The larger oncology effect is a reported interaction worth further study, not a validated indication for selective rollout. Some supplementary trigger counts and a secondary crude effect estimate do not reconcile

with their labels, as recorded below; these reporting problems warrant clarification without invalidating the coherent primary mortality counts.^{1,3}

Body of Evidence

Earlier randomised studies showed that electronic detection alone does not reliably change care. Hooper and colleagues enrolled 442 medical critical care patients meeting modified systemic inflammatory response criteria. Automated notification did not accelerate new antibiotic treatment: median times were 6.0 and 6.1 hours ($P=0.95$). Fluid administration and measured clinical outcomes also did not improve. This was a more intensively monitored setting than SCREEN's wards, and the alert offered no management recommendations. Its relevance is therefore not that all alerts fail, but that recognising an abnormal pattern electronically can add little when it does not produce a useful additional clinical action.⁴

Semler and colleagues subsequently studied an integrated electronic sepsis assessment and management tool in 407 medical and surgical critical care patients. The tool assembled clinical information, offered individualised assessments and facilitated orders. It did not significantly shorten completion of all indicated six-hour resuscitation bundle elements: hazard ratio 1.98 (95% CI 0.75–5.20; $P=0.159$). The wide interval leaves uncertainty, but the trial did not demonstrate the expected process benefit. Its comparison concerned access to decision support after clinical identification of sepsis, illustrating another intervention distinct from ward-wide detection and communication.⁵

Tarabichi and colleagues tested an emergency department warning system combined with pharmacist notification in a single-centre randomised quality-improvement study. Among 598 patients, median time from arrival to antibiotic administration was 2.3 hours with the programme and 3.0 hours with standard care ($P=0.039$). Days alive and out of hospital at 28 days also favoured the intervention. This supports the possibility that a warning tied to a specific, actionable workflow can alter care. It does not establish that the same process explains SCREEN, whose early antibiotic-ordering comparison was neutral, or that all electronic sepsis systems have equivalent effects.⁶

The 2021 Surviving Sepsis Campaign recommended performance-improvement programmes incorporating screening for acutely ill, high-risk patients. It also recommended against qSOFA as the sole screening tool compared with SIRS, NEWS or MEWS, reflecting concern about sensitivity. SCREEN does not directly compare those scores. A positive programme-level trial using qSOFA cannot determine whether another trigger would perform better, nor justify withholding assessment from a patient with

suspected sepsis whose qSOFA is below two. The choice of threshold and the effectiveness of the surrounding clinical response remain separate questions.⁷

In the accompanying editorial, published online with SCREEN in December 2024, Angus emphasised the importance of evaluating the complete clinical intervention. He also highlighted the unresolved mechanism and the potential for broader changes in ward practice. That interpretation fits the trial's main strength and limitation: it provides randomised evidence of benefit from one implementation, while leaving uncertainty about which components mattered and where the result will reproduce. Qualitative assessment of staff experience and careful description of local workflow would complement outcome data when testing similar programmes in other hospitals.²

Taken together, the evidence available by the end of 2024 supported rigorous evaluation of electronic warning programmes rather than assuming that better prediction guarantees better outcomes. SCREEN substantially strengthens that evidence base, but does not validate every sepsis alert or establish superiority of a particular score. A hospital considering a similar programme would need the capacity to obtain reliable observations, ensure clinical assessment, support escalation and monitor both useful responses and unintended treatment. Replication across different electronic records, staffing models and existing warning systems would clarify whether the mortality effect persists. Measurement should retain the whole eligible population and include patient outcomes, workload, antimicrobial decisions and treatment-related harms, rather than judging success solely by the proportion of alerts acknowledged.^{1,2}

Does SCREEN support electronic sepsis screening on hospital wards?

It supports the tested alert, training and clinical-response programme, which reduced adjusted in-hospital mortality. Its effects on dialysis, escalation and antimicrobial harm require monitoring, and the findings do not validate qSOFA alone.

References

1. Arabi YM, Alsaawi A, Alzahrani M, Al Khathaami AM, AlHazme RH, Al Mutfafy A, et al.; for the SCREEN Trial Group and the Saudi Critical Care Trials Group. Electronic sepsis screening among patients admitted to hospital wards: a stepped-wedge cluster randomized trial. *JAMA* 2025;333(9):763–773.
2. Angus DC. Do sepsis alerts help? *JAMA* 2025;333(9):759–760.

3. SCREEN Trial Group and Saudi Critical Care Trials Group. Supplement 2 to the SCREEN trial. *JAMA* 2025;333(9):763–773. Published supplementary methods and results.
4. Hooper MH, Weavind L, Wheeler AP, Martin JB, Gowda SS, Semler MW, et al. Randomized trial of automated, electronic monitoring to facilitate early detection of sepsis in the intensive care unit. *Crit Care Med* 2012;40(7):2096–2101.
5. Semler MW, Weavind L, Hooper MH, Rice TW, Gowda SS, Nadas A, et al. An electronic tool for the evaluation and treatment of sepsis in the ICU: a randomized controlled trial. *Crit Care Med* 2015;43(8):1595–1602.
6. Tarabichi Y, Cheng A, Bar-Shain D, McCrate BM, Reese LH, Emerman C, et al. Improving timeliness of antibiotic administration using a provider and pharmacist facing sepsis early warning system in the emergency department setting: a randomized controlled quality improvement initiative. *Crit Care Med* 2022;50(3):418–427.
7. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving Sepsis Campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med* 2021;47(11):1181–1247.

Possible Reporting Inconsistencies. In eTable 13, the listed qSOFA trigger components total 6,972 occurrences among 4,299 screening patients, fewer than the 8,598 required if every first alert had two recorded components. The relationship between these counts and the stated 12-hour trigger rule needs clarification. Table 3's crude ventilation risk ratio of 1.10 does not reconcile with 814/29,442 versus 700/30,613, which gives approximately 1.21. The chapter uses the reported adjusted comparison and draws no ventilation benefit. These issues do not change the reported primary mortality counts.

ACCESS

Giamarellos-Bourboulis EJ, Siampanos A, Bolanou A, Doulou S, Kakavoulis N, Tsiakos K, et al. Clarithromycin for early anti-inflammatory responses in community-acquired pneumonia in Greece (ACCESS): a randomised, double-blind, placebo-controlled trial. *Lancet Respir Med* 2024;12(4):294–304.

Introduction

Macrolides occupy an unusual position in the treatment of community-acquired pneumonia. They extend antibacterial coverage to organisms inadequately treated by β -lactams, but may also alter the host response to infection. These actions are difficult to separate clinically. Improvement after adding clarithromycin might reflect treatment of an unrecognised pathogen, a change in immune function, or both. Demonstrating that patients recover faster does not, by itself, identify which mechanism produced the benefit.

The distinction matters because additional antibiotics carry costs beyond the immediate prescription. A useful adjunct must offer enough clinical benefit to justify adverse effects, interactions and greater antimicrobial exposure. Conversely, withholding a macrolide because its antibacterial coverage appears redundant could miss a beneficial host-directed effect. Trials comparing antibiotic strategies have not provided a uniform answer, partly because they enrolled different populations and measured different outcomes.

ACCESS approached this question in patients with radiologically confirmed pneumonia, organ dysfunction and an elevated procalcitonin concentration. Its central hypothesis concerned early improvement rather than survival. Oral clarithromycin was added to usual β -lactam treatment, and success required both an improvement in respiratory symptoms and a favourable change in organ dysfunction or procalcitonin. Blood sampling provided an accompanying investigation of immune responses.¹

Although published in a respiratory journal and conducted on internal medicine wards, ACCESS is relevant to critical care because it studied a population at appreciable risk of deterioration and death. However, prevention of deterioration before ventilation or vasopressors is a different clinical task from reversing established, treatment-requiring organ failure. The trial must therefore be assessed on two levels: whether clarithromycin improved recovery in the patients actually enrolled, and how far that result supports treatment of patients already receiving intensive organ support.

Synopsis

ACCESS was a phase 3, double-blind, placebo-controlled trial conducted in 18 internal medicine departments in Greek public hospitals during 2021–2023. Adults required a new pulmonary infiltrate on chest radiography or computed tomography, compatible examination findings, and at least two of cough, purulent sputum, dyspnoea or pleuritic pain. Additional requirements were at least two systemic inflammatory response criteria, a Sequential Organ Failure Assessment (SOFA) score of 2 or more, and procalcitonin of at least 0.25 ng/mL. Important exclusions included recent healthcare exposure, prior macrolide treatment for the current episode, COVID-19, substantial immunosuppression and contraindications to clarithromycin, including marked QT prolongation or specified interacting drugs.^{1,2}

Of 979 patients screened, 278 were randomised equally between the groups. Five assigned clarithromycin and six assigned placebo subsequently withdrew consent and requested deletion of their data. The reported efficacy and safety populations therefore comprised 134 and 133 patients, respectively. Follow-up was complete for these 267 participants. Randomisation was stratified by study site, and matching tablets maintained masking of patients, treating personnel and outcome assessors.

Clarithromycin was given orally at 500 mg every 12 hours for seven days. Both groups received intravenous usual care, principally ceftriaxone or a β -lactam– β -lactamase inhibitor combination, with renal adjustment where appropriate. Additional Gram-positive coverage was permitted. Identification of *Legionella* or another atypical pathogen prompted treatment with intravenous moxifloxacin in either group. The comparison was consequently the addition of clarithromycin to a clinically adaptable antibiotic regimen, rather than a comparison in which all non-study treatment was identical.^{1,2}

Participants were predominantly elderly, with median ages of 81.5 and 81 years. All were reported as White. Median Pneumonia Severity Index scores were 122 and 121, and median SOFA scores were 3 and 4. Around one fifth had chronic obstructive pulmonary disease. None was receiving mechanical ventilation or vasopressors at baseline. These were medically complex ward patients with substantial prognostic risk, rather than patients already requiring those forms of organ support. A pathogen was identified in approximately 55% and 53%, using the study's microbiological investigations. Only four participants had documented *Legionella* infection. Piperacillin–tazobactam was used in over one third of each group, and some patients received corticosteroids during treatment.^{1,2}

The primary endpoint was assessed on day 4, after approximately 72 hours of treatment. It required a reduction of at least 50% in the respiratory symptom severity score without a need to change usual care, together with either a reduction of at least 30% in SOFA score or a favourable procalcitonin response, or both. The procalcitonin criterion was a fall of at least 80% from baseline or a concentration below 0.25 ng/mL. The respiratory score summed the four presenting symptoms, each graded from absent to severe. The trial was designed with 90% power to detect primary response rates of 60% versus 40%; it was not sized primarily for mortality.¹

The primary endpoint occurred in 91 of 134 patients receiving clarithromycin, 67.9%, compared with 51 of 133 receiving placebo, 38.3%. The absolute difference was 29.6 percentage points (95% CI 17.7–40.3), and the odds ratio was 3.40 (95% CI 2.06–5.63; $P < 0.0001$). Exploratory examination of the components showed favourable results in the same direction: respiratory symptom improvement occurred in 72% versus 48%, SOFA improvement in 68% versus 41%, and the procalcitonin response in 69% versus 54%. All three criteria were met by 42% versus 20%.¹

Clinical success, defined as complete resolution of the respiratory symptoms, was also more frequent with clarithromycin at the day 8, day 14 and day 28 assessments. Progression to organ dysfunction by day 28 occurred in 8 of 134 patients, 6%, versus 23 of 133, 17% (odds ratio 0.30; 95% CI 0.13–0.71). This endpoint concerned a new requirement for mechanical ventilation or continuous positive airway pressure, or initiation of vasopressors. A new episode of sepsis after resolution of the original pneumonia occurred in 18 versus 32 patients. These events provided clinical support for the early response finding, although they were secondary outcomes.^{1,2}

By day 90, 106 patients assigned clarithromycin and 83 assigned placebo had been discharged alive, approximately 79% versus 62%. Mortality, however, was not significantly different. At day 28, 27 of 134 patients, 20.1%, had died with clarithromycin, compared with 35 of 133, 26.3%, with placebo; the odds ratio was 0.70 (95% CI 0.39–1.25). At day 90, deaths numbered 46 and 50, 34.3% versus 37.6%, giving an odds ratio of 0.87 (95% CI 0.52–1.43). A post hoc analysis of day 8 mortality favoured clarithromycin, but the later survival comparisons remained inconclusive. Hospital readmission was not significantly reduced.¹

The biological investigation found greater ex vivo tumour necrosis factor production after lipopolysaccharide stimulation of peripheral blood mononuclear cells in the clarithromycin group on day 4. Serum interleukin-10 findings and an interleukin-8/interleukin-10 ratio criterion also favoured clarithromycin. These analyses involved 127 and 119 patients with available day 4 samples. Other cytokine comparisons

were not consistently different, and several serum measurements were below assay detection limits.^{1,2}

At least one serious treatment-emergent adverse event occurred in 58 of 134 patients receiving clarithromycin, 43%, and 70 of 133 receiving placebo, 53%. The corrected odds ratio was 0.67 (95% CI 0.42–1.11), without a statistically significant difference. Investigators attributed no serious event to study treatment. A correction published on 21 February 2024 addressed this odds ratio and a figure's risk-set count, and also corrected the appendix; the supplied versions incorporate those changes.^{1,3}

Critique

ACCESS has several features that strengthen its central result. Allocation was concealed, treatment was masked with matching placebo, and the study was conducted across multiple departments. Radiological confirmation, organ dysfunction and procalcitonin enrichment reduced the diagnostic heterogeneity often encountered in trials of suspected pneumonia. The primary difference was large and its confidence interval excluded a small effect. Concordant changes in symptoms, SOFA and procalcitonin make it unlikely that the composite result arose entirely from one uninformative component.¹

Nevertheless, the endpoint requires careful translation. A patient could succeed through symptom improvement plus either the SOFA or procalcitonin criterion. Success did not require all three to improve, complete symptom resolution, discharge or survival. The odds ratio of 3.40 describes the odds of this response and should not be interpreted as a threefold increase in the probability of recovery. The absolute response difference gives a more intuitive account of what the trial established. It remains a meaningful early recovery result, with a narrower scope than the phrase improved outcomes might imply.

Relative changes in coarse scores introduce additional sensitivity to baseline values. With a SOFA score of 3, a one-point reduction exceeds the 30% threshold; with a score of 4, the same reduction does not. Median baseline SOFA was lower in the clarithromycin group. This chance imbalance does not invalidate randomisation or explain away the large overall result, but it illustrates why percentage-based response thresholds need scrutiny. The symptom score also incorporated patient-reported and clinically judged features, making successful masking particularly valuable.¹

The analysis was not a complete intention-to-treat analysis of everyone randomised. Eleven participants were omitted after requesting data deletion. The exclusions were numerically balanced and understandable, and there was no subsequent missing follow-up among those retained. Even so, the analysed population should be reported as

267 rather than 278. Claims that all randomised patients contributed to the primary comparison would overstate the completeness of the evidence.

The secondary clinical findings are encouraging but do not all carry equal evidential weight. Fewer patients required new respiratory or circulatory support, and more achieved complete symptom resolution. These observations help establish that the main result was not solely a laboratory phenomenon. However, numerous secondary endpoints, time points, component analyses and biomarkers were examined without a clearly described multiplicity correction in the main report. Nominally significant findings, particularly those close to 0.05, should therefore support rather than independently prove the treatment claim. The individual primary-component analyses and the day 8 mortality analysis were explicitly described as post hoc.¹

Mortality is the clearest boundary to interpretation. The day 28 point estimate favoured clarithromycin, but its confidence interval included no benefit and clinically important uncertainty in both directions. The day 90 estimate was closer to the null. Lack of statistical significance does not prove the absence of a survival effect, yet inadequate power does not establish that an observed numerical reduction is real. Selecting day 8 because that comparison crossed a significance threshold would privilege an exploratory time point over the more complete survival follow-up. ACCESS demonstrated improved early response; it did not demonstrate improved 28-day or 90-day survival.

Discharge alive also needs to be considered alongside death. Patients who die can no longer experience a subsequent live discharge, while survivors may leave hospital at different times. The higher proportion discharged alive is clinically relevant, but a time-to-discharge model alone does not separate faster recovery from the competing influence of mortality. Nor does it demonstrate a reduction in total healthcare costs. Such a claim would require an economic analysis incorporating treatment, subsequent care and survival, rather than an extrapolation from the discharge curve.^{1,2}

The mechanistic interpretation is more tentative than the clinical comparison. Randomisation tested a drug with both antibacterial and immunological actions, not an immune-specific intervention against an antibiotic-matched comparator. A pathogen was unidentified in almost half the cohort, and the available microbiology did not exclude every possible antimicrobial contribution. Similar response patterns in pathogen subgroups cannot prove that bacterial coverage was irrelevant. The trial also did not establish that the measured cytokine changes mediated the improvement in symptoms or organ dysfunction.^{1,2}

The laboratory observations themselves resist a simple description as suppression of inflammation. Day 4 stimulated cellular tumour necrosis factor production was higher with clarithromycin, while some serum interleukin-10 measures were lower. Higher production between groups is not the same as an increase from each group's baseline. The interleukin-8/interleukin-10 analysis classified a fall of no more than 15% as favourable; that threshold includes preservation of the ratio and does not require an increase. Several other cytokine comparisons were neutral. These findings suggest selective changes in immune responsiveness, rather than demonstrating a uniform anti-inflammatory effect.²

Furthermore, day 4 sampling excluded some patients who had died early or been transferred to intensive care. The biological comparison therefore concerned survivors with obtainable samples after treatment had begun. Missingness related to illness trajectory can disturb the balance created by randomisation and complicate mechanistic inference. Results near assay detection limits add further uncertainty about the biological size of some reported differences. The immune data provide a plausible line of investigation, but are not a validated test for selecting patients or monitoring benefit.

Applicability is also bounded. The very elderly Greek population, extensive comorbidity and exclusion of patients with contraindications differ from many mixed critical care populations. No participant required ventilation or vasopressors at enrolment. Baseline organ dysfunction and subsequent mortality make it misleading to call the cohort simply mild pneumonia, but those features do not turn it into a trial of established ventilated pneumonia or refractory septic shock. Concurrent corticosteroid use also means ACCESS was not a study confined to patients receiving no other potentially immunomodulatory treatment.^{1,2}

Finally, the absence of attributed serious drug toxicity should not be extended to people excluded because of cardiac or pharmacological risk. The sample was too small to establish the frequency of uncommon adverse effects, and the trial did not establish the ecological consequences of widespread macrolide use. Funding came from Abbott Laboratories, with the Hellenic Institute for the Study of Sepsis acting as sponsor and retaining responsibility for study conduct and analysis; relevant author relationships were disclosed. These arrangements warrant transparency, while interpretation should remain anchored in the trial's methods and results.¹

Body of Evidence

Earlier randomised evidence helps explain why ACCESS was needed. In Garin and colleagues' 2014 open-label trial, 580 patients with moderately severe community-acquired pneumonia received a β -lactam alone or with a macrolide. At day 7, 41.2% versus 33.6% had not reached clinical stability. The 7.6-percentage-point difference was not statistically significant on a superiority test, but the upper confidence bound exceeded the trial's 8-point non-inferiority margin. Monotherapy therefore failed to establish non-inferiority; the study did not prove a mortality advantage for combination treatment. Its recovery endpoint offers a closer comparison with ACCESS than a trial concerned only with survival, although the masking and response definitions differed.⁴

CAP-START addressed antibiotic policy in a broader population. This 2015 cluster randomised crossover trial enrolled 2,283 patients with clinically suspected pneumonia on non-intensive-care wards in seven Dutch hospitals. Strategies favouring β -lactam monotherapy, β -lactam plus macrolide, or fluoroquinolone monotherapy were rotated between periods, with deviations permitted for medical reasons. Using a 3-percentage-point margin, the β -lactam strategy established non-inferiority for 90-day mortality. This was evidence about preferred empirical strategies, rather than an individually randomised comparison of obligatory monotherapy with an identical regimen plus placebo. The broader diagnostic entry criteria and different primary endpoint limit a direct claim that CAP-START contradicts ACCESS.⁵

A related question in patients with established organ dysfunction was examined by INCLASS in 2022. This double-blind trial randomised 110 patients with sepsis, respiratory dysfunction and additional organ dysfunction to intravenous clarithromycin or placebo. Day 28 mortality was 49.1% with clarithromycin and 45.5% with placebo; the absolute difference was 3.6 percentage points (95% CI -15.7 to 22.7). The study did not demonstrate a survival benefit. Although small and clinically different from ACCESS, it cautions against transferring a positive early-response result in ward pneumonia directly to patients with advanced organ failure.⁶

The 2023 ERS/ESICM/ESCMID/ALAT severe pneumonia guideline suggested adding a macrolide rather than a fluoroquinolone to β -lactam treatment. That was a conditional recommendation supported by very low-quality evidence. Its comparison matters: preference between two adjunctive antibiotic classes does not settle whether every patient needs an adjunct, nor does it specifically validate seven days of oral clarithromycin for immunomodulation. ACCESS adds direct placebo-controlled evidence in an enriched ward population, while leaving the guideline's more severe population and alternative regimens incompletely addressed.⁷

The accompanying editorial interpreted ACCESS as strengthening the case for macrolides in hospitalised pneumonia and emphasised possible effects on immune function. Subsequent correspondence focused on which patients best matched the trial. The authors' reply highlighted the absence of baseline ventilation or vasopressor treatment despite appreciable pneumonia severity. This exchange usefully separates physiological and prognostic severity from the immediate requirement for organ support. It does not identify a validated subgroup in whom survival benefit is certain, and editorial enthusiasm should not replace the uncertainty of the mortality estimates.^{8,9,10}

Taken together, the evidence available by the end of 2024 supports a specific advance. ACCESS provides persuasive evidence that adding clarithromycin improves its defined early clinical and biological response in selected hospitalised patients with pneumonia. Earlier strategy trials and the neutral survival comparison in INCLASS prevent that finding from becoming a universal claim about macrolides across all pneumonia and sepsis. The remaining questions concern patient selection, the contribution of antibacterial versus immune effects, the best drug and duration, and whether improved early recovery translates into durable patient benefit. For critical care practice, the result is most directly relevant before the onset of ventilation or vasopressor dependence, with prescribing still requiring attention to microbiology, interactions and cardiac risk.

Does adjunctive clarithromycin improve outcomes in hospitalised community-acquired pneumonia?

In selected ward patients with organ dysfunction and raised procalcitonin, clarithromycin improved early response and reduced subsequent deterioration, but did not establish a survival benefit or effectiveness in patients already requiring ventilation or vasopressors.

References

1. Giamarellos-Bourboulis EJ, Siampanos A, Bolanou A, Doulou S, Kakavoulis N, Tsiakos K, et al. Clarithromycin for early anti-inflammatory responses in community-acquired pneumonia in Greece (ACCESS): a randomised, double-blind, placebo-controlled trial. *Lancet Respir Med* 2024;12(4):294–304.
2. ACCESS investigators. Supplementary appendix to the ACCESS trial. *Lancet Respir Med* 2024;12(4):294–304. Published supplementary methods and results, corrected 21 February 2024.
3. Correction to *Lancet Respir Med* 2024; published online Jan 3. [https://doi.org/10.1016/S2213-2600\(23\)00412-5](https://doi.org/10.1016/S2213-2600(23)00412-5). *Lancet Respir Med* 2024;12(4):e20.

4. Garin N, Genné D, Carballo S, Chuard C, Eich G, Hugli O, et al. β -Lactam monotherapy vs β -lactam–macrolide combination treatment in moderately severe community-acquired pneumonia: a randomized noninferiority trial. *JAMA Intern Med* 2014;174(12):1894–1901.
5. Postma DF, van Werkhoven CH, van Elden LJ, Thijsen SF, Hoepelman AI, Kluytmans JA, et al.; CAP-START Study Group. Antibiotic treatment strategies for community-acquired pneumonia in adults. *N Engl J Med* 2015;372(14):1312–1323.
6. Karakike E, Scicluna BP, Roumpoutsou M, Mitrou I, Karampela N, Karageorgos A, et al. Effect of intravenous clarithromycin in patients with sepsis, respiratory and multiple organ dysfunction syndrome: a randomized clinical trial. *Crit Care* 2022;26:183.
7. Martin-Loeches I, Torres A, Nagavci B, Aliberti S, Antonelli M, Bassetti M, et al. ERS/ESICM/ESCMID/ALAT guidelines for the management of severe community-acquired pneumonia. *Eur Respir J* 2023;61(4):2200735.
8. Waterer G. Macrolides in community-acquired pneumonia. *Lancet Respir Med* 2024;12(4):260–262.
9. Xu Y, Cai Y. Community-acquired pneumonia: the best candidates for clarithromycin. *Lancet Respir Med* 2024;12(4):e16.
10. Giamarellos-Bourboulis EJ, Niederman MS. Community-acquired pneumonia: the best candidates for clarithromycin—Authors' reply. *Lancet Respir Med* 2024;12(4):e17.

Possible Reporting Inconsistencies. Table 2 gives a 95% confidence interval of 0.4–49.6 percentage points for the 12.3-point difference in day 28 clinical success; its unusually wide upper limit warrants clarification against the underlying counts of 83/134 and 66/133. Supplementary Table 1 labels procalcitonin in pg/mL while the main report uses ng/mL for the corresponding baseline values. The chapter uses the main report's units and does not reproduce the questionable clinical-success interval.



Trauma Trials

TRAUMOX2

Arleth T, Baekgaard J, Siersma V, Creutzburg A, Dinesen F, Rosenkrantz O, et al.; TRAUMOX2 Trial Group. Early restrictive vs liberal oxygen for trauma patients: the TRAUMOX2 randomized clinical trial. JAMA 2025;333(6):479–489.

Introduction

Oxygen is commonly given early after major trauma, often before the full extent of injury is known. The immediate rationale is compelling: hypoxaemia can aggravate brain injury and compromise already threatened organs. However, giving oxygen to a patient with inadequate oxygenation and routinely exposing a patient with normal oxygenation to a high inspired concentration are different interventions. The appropriate dose, target and duration of oxygen during the initial response to trauma have remained uncertain.

There are plausible disadvantages to excessive oxygen exposure, including oxidative injury and absorption atelectasis. Yet physiological plausibility and observational associations do not establish that restricting oxygen improves patient outcomes. Injury severity, intubation, shock and clinicians' concern may all influence the amount of oxygen delivered. Patients receiving the most oxygen can therefore be the patients most likely to die for reasons unrelated to that oxygen.

Trials conducted in intensive care provide only partial guidance. They often begin after initial resuscitation, enrol populations dominated by medical illness, and maintain oxygen targets for days. The first hours after injury involve different conditions: incomplete diagnoses, transport, rapidly changing physiology and limited access to arterial blood gases. An oxygen strategy for this interval must work across ambulances, resuscitation rooms, theatres and wards.

TRAUMOX2 tested an eight-hour intervention beginning either before hospital arrival or on admission to a trauma centre. It compared titration towards a peripheral oxygen saturation of 94% with a deliberately liberal oxygen regimen. Its primary question was whether restrictive treatment reduced death or major respiratory complications within 30 days. The trial therefore addresses a clinically recognisable early management decision, while requiring care not to extend its findings to every oxygen target or the entire subsequent critical care admission.¹

Synopsis

TRAUMOX2 was an investigator-initiated, international, multicentre, open-label superiority trial with masked primary outcome assessment. Recruitment occurred between December 2021 and September 2023 at 15 prehospital bases and five major trauma centres in Denmark, the Netherlands and Switzerland. Eligible patients were adults with blunt or penetrating trauma who were transported directly to a participating trauma centre, triggered full trauma team activation, and were expected by the enrolling physician to remain in hospital for at least 24 hours. Suspected carbon monoxide intoxication and cardiac arrest before randomisation were exclusions.¹

Allocation used sealed envelopes containing a computer-generated sequence, with variable block sizes and stratification by inclusion site and intubation status. Envelopes allowed enrolment when internet or telephone coverage was unreliable. Treating clinicians and patients knew the allocation. Subsequent primary outcome adjudication used records from which information revealing the oxygen intervention had been concealed. Two medical specialists assessed each case independently, resolving disagreement through discussion or a third assessor.¹

A total of 1,979 patients were randomised, but 471 were subsequently excluded. This included patients found to have no or minor injuries and an expected stay below 24 hours after secondary survey, other eligibility problems and two cases with lost randomisation envelopes. The resulting study population comprised 750 patients allocated restrictive oxygen and 758 allocated liberal oxygen. During follow-up, outcome data were unavailable for another 17 and 34 patients, mainly because of consent withdrawal. The primary comparison consequently included 733 and 724 patients, rather than everyone originally randomised.^{1,2}

Median age was approximately 50 years, 73% were male, and about 89% had blunt injuries. Median Injury Severity Score was 14 in each group. Approximately one quarter were intubated at randomisation and 41% were enrolled before hospital arrival. Serious head and thoracic injuries were common, but the trial was not confined to patients with profound shock or those destined for intensive care. Denmark contributed most participants. Active smoking was more frequently recorded in the liberal group, while the main measured baseline characteristics were otherwise broadly similar.^{1,2}

The restrictive strategy used the lowest oxygen dose needed to maintain SpO₂ at 94%. Patients whose saturation was already at least 94% on room air did not require supplemental oxygen; there was no attempt to lower their naturally higher saturation. Oxygen could be delivered by nasal cannula, non-rebreather mask or mechanical ventilation. The liberal group received 15 L/min through a non-rebreather mask if not

intubated, or an FiO_2 of 1.0 if intubated, during prehospital care, trauma resuscitation and intrahospital transport. In theatre, recovery, intensive care or the ward, this could be reduced to 12 L/min or an FiO_2 of at least 0.6 when SpO_2 was at least 98%. Brief higher-dose oxygen, including preoxygenation for intubation, was permitted in either group.¹

The intervention lasted eight hours. Randomisation occurred a median of 54 and 52 minutes after injury among patients with timing data. Nearly half had already received supplemental oxygen. During treatment, the median oxygen flow in non-intubated patients was 0 L/min with restriction and 12 L/min with liberal treatment. Among intubated patients, median FiO_2 was 0.28 versus 0.60. Median PaO_2 at the first scheduled arterial sample was 85 versus 280 mm Hg, and at the second was 86 versus 230 mm Hg. Thus, both delivered oxygen and measured arterial oxygenation separated substantially.^{1,2}

The primary endpoint combined death with major respiratory complications, defined as pneumonia using Centers for Disease Control and Prevention criteria or acute respiratory distress syndrome using the Berlin definition. Respiratory assessment covered the available hospital record up to discharge or day 30; pneumonia after discharge was a separate exploratory outcome. Mortality was sought through records, registries and, where required, contact with patients or relatives. The sample size calculation targeted a reduction in the primary event rate from 15% to 10%, with 80% power.^{1,2}

The main analysis used logistic regression adjusted for the stratification variables, with site clustering addressed through generalised estimating equations. Inverse probability weighting addressed differential availability of follow-up data. Additional models adjusted for age, sex, injury severity and initial Glasgow Coma Scale, with a further analysis incorporating smoking. The primary outcome occurred in 118 of 733 patients receiving restrictive oxygen, 16.1%, and 121 of 724 receiving liberal oxygen, 16.7%. The reported adjusted odds ratio was 1.01 (95% CI 0.75–1.37; $P=0.94$). Additional adjustment and the per-protocol analysis did not establish a difference.^{1,2}

The two components moved numerically in different directions. Death occurred in 63 of 733 patients, 8.6%, versus 53 of 724, 7.3%, with an odds ratio of 1.28 (95% CI 0.85–1.92). Major respiratory complications occurred in 65 patients, 8.9%, versus 78, 10.8%, with an odds ratio of 0.84 (95% CI 0.59–1.19). Neither comparison was statistically significant. Subgroup analyses did not identify convincing treatment-effect modification, including among patients intubated at enrolment or those with moderate or severe traumatic brain injury.¹

At least one hourly SpO₂ recording below 90% occurred in 44 of 737 restrictive-group patients, 6.0%, and 28 of 737 liberal-group patients, 3.8%. The nominal odds ratio was 1.67 (95% CI 1.02–2.70), but no exploratory outcome remained significant after adjustment for multiple testing. Radiologically identified atelectasis was recorded in 207 of 750 patients, 27.6%, versus 263 of 758, 34.7%. Other recorded adverse-event categories were broadly similar. Major oxygenation violations were documented in approximately 7% and 14%, respectively, despite the substantial overall separation in exposure.^{1,2}

Critique

The main strength of TRAUMOX2 is its direct examination of a difficult and common decision. Recruitment extended into prehospital practice and continued across several national systems. The intervention was simple enough to follow during transport and transfer between clinical locations. Importantly, the trial achieved a large difference in oxygen dose and PaO₂. A neutral result in this setting cannot reasonably be dismissed on the grounds that the two groups received almost the same oxygen treatment.¹

The treatment labels nevertheless need precision. Restrictive oxygen meant titration towards 94%, with room air permitted when sufficient. It did not mean deliberately maintaining a saturation of 88–92%, withholding oxygen during hypoxaemia, or prohibiting preoxygenation. Indeed, median SpO₂ during the intervention was 97% in the restrictive group and 100% in the liberal group. Many patients could maintain saturations above the target without supplemental oxygen. This distinction is central to applying the trial: its conservative arm largely avoided unnecessary oxygen while preserving normoxaemia.²

The liberal comparator was also specific. High flows or inspired fractions were required even when oxygen saturation was already adequate. This produced pronounced hyperoxaemia, especially early after enrolment. Such a comparator provides a strong test of the hypothesis that early excess oxygen causes substantial clinical harm, but may be more liberal than contemporary care in some settings. A centre already titrating oxygen to adequate saturation would not be making the same change as a centre routinely using these high doses. The trial does not identify an optimal point between the two strategies.

Masking of outcome assessment was a substantial safeguard. Pneumonia and acute respiratory distress syndrome require interpretation of clinical and imaging data, so an entirely unblinded assessment could favour the strategy an investigator expected to be safer. Concealment of oxygen-related information worked reasonably well: assessors

guessed allocation correctly in about half of cases. However, clinicians remained unblinded and could influence investigations, intubation, ventilation and other care before adjudication. Masking the later diagnosis cannot remove differences in the underlying care or availability of diagnostic information.¹

The most important limitation is the large reduction from the randomised population to the analysed population. Nearly one quarter of those randomised were excluded before formation of the 1,508-patient cohort. Early trauma assessment necessarily involves uncertainty, and excluding patients subsequently found to have trivial injuries concentrates the study on its intended clinical population. Nevertheless, an assessment of likely discharge made after allocation can create selection bias if allocation influences that judgement. Similar exclusion counts are reassuring only to a degree; they do not restore the protection of including everyone as randomised. The result should be described as a modified intention-to-treat comparison.^{1,2}

Subsequent missingness was smaller but unequal. Primary outcome data were unavailable for 17 restrictive-group and 34 liberal-group patients. Weighting is a reasonable attempt to address this, but depends on the measured predictors adequately explaining the probability of observation. It cannot guarantee removal of bias from unmeasured differences between those observed and those missing. The appendix's best/worst-case table assigns all missing outcomes as non-events in both groups or as events in both groups. These are useful sensitivity scenarios, but they are not the most adverse between-group scenarios, which would assign opposite outcomes to missing patients in the two groups.²

The composite endpoint combines outcomes of very different importance. Avoiding death and avoiding pneumonia cannot be treated as interchangeable benefits, and a patient may experience both. Mortality was numerically higher with restriction, while major respiratory complications were numerically lower. These uncertain trends do not establish a trade-off, but they mean that a neutral composite should not be read as reassuring evidence of benefit in every component. The separate mortality confidence interval remains compatible with both benefit and appreciable harm.

TRAUMOX2 was a superiority trial. Failure to show that restrictive treatment was better does not establish non-inferiority or equivalence. Its size was chosen to detect a five-percentage-point absolute reduction, a large effect for a short intervention. Smaller differences could still matter because oxygen is used so widely. The primary confidence interval is therefore more informative than the near-unity point estimate alone: the study does not demonstrate a clinically important advantage, but it does not establish that the strategies have identical consequences.¹

The eight-hour window is both an advantage and a limitation. It isolates the period in which empirical high-dose oxygen is commonly given, without attempting to control all later care. However, roughly half the patients had already received oxygen before randomisation, and the subsequent hospital course was not governed by the allocated strategy. Early exposure before enrolment or later exposure after eight hours could attenuate a treatment effect. The findings are most directly about changing oxygen treatment during this particular interval, rather than the consequences of sustained hyperoxaemia throughout critical illness.^{1,2}

Outcome surveillance was also bounded. The respiratory component used available inpatient information and did not incorporate all postdischarge respiratory events into the primary endpoint. Transfer to a hospital with inaccessible records could shorten the available assessment. Postdischarge pneumonia was evaluated separately and by methods that varied between sites. Death ascertainment was more extensive, but the supplementary methods allowed patients discharged home and not contactable after repeated attempts to be considered alive. These details qualify a simple assertion that every participant underwent comprehensive 30-day clinical surveillance.²

The hypoxaemia and atelectasis observations merit attention without displacing the primary result. Hourly saturation recordings identified more low-saturation episodes with restriction, but could not measure the full duration or burden of transient hypoxaemia. That exploratory difference did not survive multiplicity adjustment. Less atelectasis with restriction is consistent with a plausible oxygen-related effect, yet the event was a radiological finding and was not synonymous with improved respiratory function or fewer major respiratory complications. Adverse-event denominators also included patients observed only until consent withdrawal; absence of a recorded event did not necessarily mean complete 30-day surveillance.²

Finally, the enrolled population included many patients who were never profoundly hypoxaemic or critically unstable. Findings in the entire cohort cannot establish safety in every patient with severe brain injury, major haemorrhage or chronic respiratory disease. Subgroup confidence intervals were wide, and subgroup membership based on subsequent intensive care admission is itself determined after randomisation. Race and ethnicity were not collected, limiting evaluation of performance across skin pigmentation when treatment depended on pulse oximetry. These limitations support careful titration and reassessment, rather than an inflexible saturation rule for every trauma patient.¹

Body of Evidence

TRAUMOX1 established feasibility before the larger trial. It randomised 41 patients to 24 hours of restrictive or liberal oxygen after trauma. Arterial oxygenation separated substantially, and the composite of death or major inpatient pulmonary complications occurred in 4 of 20 versus 6 of 18 evaluable patients. Two patients died in each group. These small numbers supported the practicality of maintaining normoxaemia with less oxygen, but could not establish efficacy. TRAUMOX2 supplied a much larger test and used a shorter intervention extending into prehospital care.³

The 2018 IOTA meta-analysis contributed to concern about liberal oxygen. Across 25 trials involving 16,037 acutely ill adults, liberal treatment was associated with higher hospital mortality (relative risk 1.21; 95% CI 1.03–1.43). The included illnesses and oxygen regimens were heterogeneous. This finding strengthened the rationale for avoiding unexamined high-dose oxygen, but it could not directly determine the effect of an eight-hour strategy in trauma. A broad pooled estimate should not substitute for the later population-specific randomised comparison.⁴

ICU-ROX subsequently randomised 1,000 mechanically ventilated adults to conservative or usual oxygen therapy. The conservative strategy reduced oxygen exposure but did not significantly improve ventilator-free days: the reported absolute difference was –0.3 days (95% CI –2.1 to 1.6). Its usual-care comparator did not mandate the high oxygen concentrations used in TRAUMOX2. The two trials therefore examined different contrasts, despite both using the language of conservative oxygen.⁵

HOT-ICU studied 2,928 patients with acute hypoxaemic respiratory failure, comparing PaO₂ targets of 60 and 90 mm Hg. Ninety-day mortality was 42.9% versus 42.4%, without a significant difference. Its participants required treatment for established respiratory failure, and both the intervention and comparator differed substantially from the early trauma regimens. It does not establish a universal target for injured patients.⁶

PILOT compared SpO₂ targets of 90%, 94% and 98% in 2,541 mechanically ventilated adults. Median ventilator-free days were 20, 21 and 21, without a significant difference. Although the intermediate target resembles TRAUMOX2's restrictive target numerically, the clinical setting, duration and alternatives were different. Identical target numbers do not make the tested strategies interchangeable.⁷

LOCO2 provides a caution against interpreting oxygen restriction as invariably harmless. In patients with acute respiratory distress syndrome, it compared a conservative target of PaO₂ 55–70 mm Hg or SpO₂ 88–92% with a higher target. The trial stopped after 205 enrolments because of safety concerns and low likelihood of demonstrating its primary

benefit. Day 28 mortality did not differ significantly, while later mortality and mesenteric ischaemia raised concern. These lower oxygenation targets were not those tested in TRAUMOX2, but the experience reinforces the need to avoid extrapolating from reduced excess oxygen to permissive hypoxaemia.⁸

In 2024, HOT-COVID provided a different result in a specific disease population. Among 726 randomised patients with severe COVID-19, targeting PaO₂ at 60 rather than 90 mm Hg increased days alive without life support, without establishing a survival benefit. The intervention continued during intensive care rather than ending after eight hours. This finding illustrates why the illness, treatment interval and outcome must remain attached to any conclusion about the preferred oxygen strategy.⁹

The evidence available at the end of 2024 therefore resists a single rule that more or less oxygen is always better. TRAUMOX2 adds direct evidence for the initial hours after injury, demonstrating that a substantial reduction in oxygen exposure did not produce the large improvement in death or major respiratory complications that the investigators sought. It also fails to establish that compulsory high-dose oxygen improves those outcomes. Its practical contribution is to make routine empirical oxygen dosing open to scrutiny while preserving rapid treatment of hypoxaemia. Decisions still need to account for the patient's oxygenation, injury pattern, perfusion and trajectory, and for the limited precision of evidence in the most vulnerable subgroups.

Does restrictive oxygen during the first eight hours of trauma care improve clinical outcomes?

Targeting SpO₂ at 94% reduced oxygen exposure but did not significantly reduce death or major respiratory complications. Equivalence was not established, and the trial does not support withholding oxygen from hypoxaemic patients.

References

1. Arleth T, Baekgaard J, Siersma V, Creutzburg A, Dinesen F, Rosenkrantz O, et al.; TRAUMOX2 Trial Group. Early restrictive vs liberal oxygen for trauma patients: the TRAUMOX2 randomized clinical trial. *JAMA* 2025;333(6):479–489.
2. TRAUMOX2 Trial Group. Supplemental online content to the TRAUMOX2 randomised clinical trial. *JAMA* 2025;333(6):479–489. Published supplementary methods and results.

3. Baekgaard JS, Isbye D, Ottosen CI, Larsen MH, Andersen JH, Rasmussen LS, et al. Restrictive vs liberal oxygen for trauma patients—the TRAUMOX1 pilot randomised clinical trial. *Acta Anaesthesiol Scand* 2019;63(7):947–955.
4. Chu DK, Kim LH, Young PJ, Zamiri N, Almenawer SA, Jaeschke R, et al. Mortality and morbidity in acutely ill adults treated with liberal versus conservative oxygen therapy (IOTA): a systematic review and meta-analysis. *Lancet* 2018;391(10131):1693–1705.
5. ICU-ROX Investigators and the Australian and New Zealand Intensive Care Society Clinical Trials Group. Conservative oxygen therapy during mechanical ventilation in the ICU. *N Engl J Med* 2020;382(11):989–998.
6. Schjørring OL, Klitgaard TL, Perner A, Wetterslev J, Lange T, Siegemund M, et al.; HOT-ICU Investigators. Lower or higher oxygenation targets for acute hypoxemic respiratory failure. *N Engl J Med* 2021;384(14):1301–1311.
7. Semler MW, Casey JD, Lloyd BD, Hastings PG, Hays MA, Stollings JL, et al.; PILOT Investigators and the Pragmatic Critical Care Research Group. Oxygen-saturation targets for critically ill adults receiving mechanical ventilation. *N Engl J Med* 2022;387(19):1759–1769.
8. Barrot L, Asfar P, Mauny F, Winiszewski H, Montini F, Badie J, et al.; LOCO2 Investigators and REVA Research Network. Liberal or conservative oxygen therapy for acute respiratory distress syndrome. *N Engl J Med* 2020;382(11):999–1008.
9. Nielsen FM, Klitgaard TL, Siegemund M, Laake JH, Thormar KM, Cole JM, et al.; HOT-COVID Trial Group. Lower vs higher oxygenation target and days alive without life support in COVID-19: the HOT-COVID randomized clinical trial. *JAMA* 2024;331(14):1185–1194.



Paediatric Trials

VODE

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.

Introduction

Intubating a newborn infant combines a technically demanding procedure with limited physiological reserve. The airway is small, the margin for incorrect tube placement is narrow, and repeated attempts can interrupt effective ventilation. The procedure also has an important educational dimension: in many neonatal units, doctors in training make the first attempt under senior supervision. The equipment must therefore support both the immediate task and the communication needed to perform it safely.

Video laryngoscopy changes what the operator can see and what other members of the team can share. A camera close to the blade tip can provide an indirect view of the larynx, while a screen may allow a supervisor to recognise difficulty and guide the operator. However, improved visualisation does not necessarily translate into easier tube passage, quicker intubation or fewer complications. Different devices, blade sizes, routes of intubation and levels of experience may produce different results.

Evidence from adult emergency intubation or elective anaesthesia in older infants cannot settle the neonatal question. Urgent intubation in the delivery room often occurs before intravenous access is available, while intubation in neonatal intensive care commonly follows non-invasive respiratory support and premedication. Extremely preterm infants introduce further concerns about blade size, airway trauma and rapid desaturation.

VODE examined video laryngoscopy as the operator's method of viewing the airway during urgent oral intubation. It included newborn infants across gestational ages and enrolled in both the delivery room and neonatal intensive care unit. The primary endpoint was successful first-attempt intubation confirmed by exhaled carbon dioxide. The key question is whether a large improvement in that procedural outcome justifies a change in equipment and training, while recognising that evidence of fewer attempts is not automatically evidence of less physiological harm.¹

Synopsis

VODE was a single-centre randomised trial at the National Maternity Hospital in Dublin, a university maternity hospital with approximately 7,000 births annually. Recruitment began in September 2021, and consent for the final infant was obtained in November 2023. Newborn infants of any gestational age undergoing intubation in the delivery room or neonatal intensive care unit were eligible. Upper airway anomalies were excluded. Each infant could enter only once, although an infant previously intubated at another hospital could participate if subsequently intubated at the study centre.¹

Of 246 infants assessed, 241 were considered eligible and 226 were randomised. Fifteen eligible infants were not randomised, chiefly because a researcher was unavailable. Allocation assigned 115 infants to video laryngoscopy and 111 to direct laryngoscopy. One video-group infant improved and did not undergo intubation; deferred consent was not obtained for another seven video-group and four direct-group infants. The primary analysis therefore included 107 infants in each group, representing 89% of eligible infants.¹

Randomisation was stratified by gestational age below or at least 32 weeks, using sequentially numbered, sealed opaque envelopes. The senior clinician selected the operator for the first attempt before allocation was revealed. Multiple-birth infants were randomised individually. Prospective parental consent was sought where practical, with an approved deferred consent process for urgent circumstances. Clinicians, researchers recording the procedure and analysts knew the allocation. The radiologist assessing final endotracheal tube position was masked.¹

The intervention used a C-MAC video laryngoscope with a Miller blade and an eight-inch monitor. Size 0 was used below 1,500 g and size 1 at higher weights. The comparator used a conventional HEINE laryngoscope for a direct view through the mouth, with straight blades of sizes 00, 0 or 1 according to weight. The blade-size thresholds therefore differed between devices. Both used uncuffed oral endotracheal tubes, and stylets were used during nearly all first attempts. Clinicians starting their rotations received instruction in both techniques and were encouraged to practise on manikins.¹

Most first attempts were made by doctors training in paediatrics or neonatology. Trainees could make up to three attempts, while a neonatologist could intervene at any point. Premedication was generally given in neonatal intensive care unless urgency prevented it, usually comprising fentanyl, atropine and suxamethonium. Infants intubated in the delivery room did not receive these drugs because intravenous access was not immediately available. Neither group received supplemental oxygen during the intubation attempt. Other procedural decisions were left to the treating team.¹

The first attempt was the allocated intervention. Crossover was strongly discouraged and was permitted after a failed attempt by a neonatologist or in exceptional circumstances. There was one initial allocation violation: an infant assigned video laryngoscopy underwent a first attempt with direct laryngoscopy because the video device was not immediately available. That infant remained in the assigned group for analysis. A researcher attended procedures and recorded attempts, their duration and observed physiological values contemporaneously.¹

An attempt began when the laryngoscope entered the mouth with the intention of placing an endotracheal tube, whether or not a tube was actually introduced. Successful intubation required a colour change on an exhaled carbon dioxide detector over six breaths. Secondary outcomes included oxygen saturation, heart rate, number of attempts, duration of successful attempts, crossover and radiographic tube position. The sample size was based on increasing first-attempt success from 40% to 60%, with 80% power, and was increased from 194 to 214 to allow for attrition. The primary comparison used a chi-square test.¹

Approximately two thirds of infants were born before 32 weeks, 43% before 28 weeks, and 42% weighed below 1,000 g at birth. Most were being intubated for the first time and more than four fifths underwent the procedure within 24 hours of birth. Overall, 63 infants were intubated in the delivery room and 151 in neonatal intensive care. Delivery-room intubation was more common in the video group, 38 versus 25 infants, and premedication was correspondingly less frequent, 63% versus 75%.¹

First-attempt intubation succeeded in 79 of 107 infants with video laryngoscopy, 74% (95% CI 66–82), compared with 48 of 107 with direct laryngoscopy, 45% (95% CI 35–54; $P < 0.001$). This represents approximately 29 additional first-attempt successes per 100 infants. The median number of attempts required was one versus two. Post hoc subgroups favoured video laryngoscopy in both locations and among the smallest infants, but were not adjusted for multiple comparisons.¹

Successful first attempts lasted a median of 61 seconds with video laryngoscopy and 51 seconds with direct laryngoscopy. A post hoc analysis found median cumulative laryngoscopy time to successful intubation of 66 versus 96 seconds. The median lowest oxygen saturation during the first attempt was 74% versus 68%, based on 98 and 100 infants with measurements. Median lowest heart rate was 153 versus 148 beats/min, with data for 99 infants per group. These were descriptive secondary comparisons with unadjusted confidence intervals.^{1,2}

Crossover was reported in 3% versus 29% among 99 infants assessed in each group. The tube tip was in the study's specified radiographic position in 53% versus 50% of 105 infants assessed per group. Correct position meant a tip between the upper border of the first and lower border of the second thoracic vertebra. Severe desaturation still occurred frequently, and the trial did not establish a reduction in serious complications. Three infants died during the trial period, one in the video group and two in the direct group; investigators considered none of these deaths related to intubation.¹

Critique

VODE provides a clear answer to its primary question. The difference in first-attempt success was large, exceeded the effect anticipated in the sample size calculation, and was unlikely to be a chance finding under the reported analysis. Allocation was concealed until the decision to intubate and the initial operator had been selected. This is an important protection against assigning a more experienced clinician after learning which device would be used. High recruitment of eligible infants, inclusion of very small newborns and contemporaneous recording strengthen the result.¹

The outcome is also clinically relevant. Avoiding repeated laryngoscopy reduces the number of times the procedure must be restarted and may simplify urgent airway management. Confirmation with exhaled carbon dioxide makes the definition more objective than an operator's impression of successful tube placement. Nevertheless, first-attempt success is a procedural endpoint. It is not synonymous with adequate tube depth, freedom from hypoxaemia, avoidance of injury or long-term neurological benefit. Those outcomes require separate evidence.

The unit of an attempt deserves attention. A new attempt required another introduction of the laryngoscope; several efforts to pass the tube while the blade remained in the mouth could still form part of one attempt. The investigators did not record the number of tube insertion efforts, the grade of laryngeal view or reasons for abandoning laryngoscopy. The trial therefore establishes fewer unsuccessful laryngoscope insertions, but cannot fully explain whether the benefit arose from better visualisation, better tube guidance, more effective supervision or a combination.¹

This was also a comparison of two equipment systems rather than the addition of a camera to an otherwise identical blade. Blade sizes and weight thresholds differed, and only one video device was studied. The result is most directly applicable to this C-MAC configuration and the training provided with it. It cannot automatically be transferred to every video laryngoscope, especially devices with different blade geometry or

performance in the smallest infants. Availability and familiarity remain part of the intervention.

The trial was unblinded at almost every stage. Researchers observed procedures in real time, and clinicians knew the assigned device. They could differ in how long they persisted, how they positioned the infant or tube, and when they sought help. Such differences may be part of the practical advantage of a video system, rather than simply bias, but they prevent isolation of a purely optical effect. A shared screen could facilitate coaching; the study did not quantify the amount or content of that coaching.¹

Operator experience is particularly relevant. Forty-five clinicians made first attempts, with most procedures undertaken by trainees. This supports the relevance of the trial to a teaching neonatal unit, but the number of previous direct or video intubations performed by each clinician was not recorded. The primary chi-square analysis did not explicitly model repeated procedures by the same operator. Experience, learning over time and clinician-specific performance could therefore affect precision and the size of benefit expected elsewhere. Results from the small number of consultant first attempts cannot establish how much an already expert operator would gain.^{1,2}

Chance imbalances also warrant context. More video-group infants were intubated in the delivery room and fewer received premedication, while the number of first attempts made by a neonatologist was higher in that group. These factors do not all favour the same interpretation. The study was not stratified by setting or clinician role, and its primary analysis did not adjust for them. Post hoc results by role and location remained encouraging, but cannot replace a sufficiently powered assessment of treatment-effect differences. In particular, large numerical benefits below 1,000 g should not be interpreted as proof that birth weight modifies efficacy.^{1,2}

The time results require a different reading from the success results. A successful first attempt took longer in the video group by the reported medians. This is a comparison restricted to successful attempts, and the intervention substantially changed which infants entered that subset. It does not show that assigning video laryngoscopy delays the overall procedure by ten seconds. Conversely, the shorter cumulative laryngoscopy time in the post hoc analysis did not measure the full elapsed time from preparation to restored ventilation, including pauses and other care. Neither result supports a simple claim that every aspect of intubation becomes faster.^{1,2}

Physiological instability remained common. Among infants with recorded values, oxygen saturation fell below 70% in approximately 45% of the video group and 53% of the direct group. Bradycardia, chest compressions and other adverse events were too infrequent or

imprecisely estimated to establish a reliable treatment difference. Oxygen and heart-rate measurements were unavailable for some infants, and multiple thresholds and complications were added post hoc. The authors explicitly cautioned against treating the unadjusted confidence intervals as confirmatory hypothesis tests. The trial's positive success endpoint should therefore not be promoted into proof of fewer serious adverse events.¹

The absence of oxygen delivery during laryngoscopy is an important boundary. It was common to both groups and does not invalidate their comparison, but it means VODE did not test video laryngoscopy within an intubation strategy that routinely maintained nasal oxygen during the attempt. Combining the two interventions may be attractive, yet their benefits cannot be assumed simply to add together. The duration of an attempt, the criteria for pausing and the use of oxygen are interdependent clinical decisions.

Later outcomes were also affected by rescue treatment. Greater crossover from direct to video laryngoscopy helps explain how initial assignment translated into the overall procedure, but subsequent attempts no longer represented an uncontaminated device comparison. Crossover estimates used fewer infants than the primary analysis, and the report does not fully explain that denominator. Radiographic assessment was stronger because it was masked, but correct tube depth remained close to one half in both groups under the study's definition. Improved entry into the trachea did not solve the separate problem of positioning the tube appropriately.¹

Finally, twelve randomised infants did not enter the primary analysis, principally because deferred consent was not obtained. The remaining infants were analysed as assigned, but this was not complete follow-up of all 226 randomised infants. Generalisation is further limited by one Irish hospital, exclusion of airway anomalies and the resources required to purchase, maintain and reprocess the equipment. The hospital foundation funded the trial, and the manufacturer loaned additional units and offered a discounted purchase after it ended, while having no reported role in study design or analysis. These practical considerations accompany a persuasive procedural benefit rather than negating it.¹

Body of Evidence

Earlier neonatal trials investigated different uses of a video system. In O'Shea and colleagues' 2015 trial, inexperienced operators performed direct laryngoscopy while the supervising clinician either could or could not see the video screen. First-attempt success was 69 of 104 intubations, 66%, with screen access and 42 of 102, 41%, without it. The study demonstrated the value of shared visual information for supervision. VODE examined the operator's use of the video view itself, so the interventions should not be treated as identical.³

VISI extended the evidence to anaesthetised infants. This multicentre trial randomised 564 infants in operating theatres, with 552 included in the main analysis. First-attempt success was 93% with video and 88% with direct laryngoscopy, an adjusted absolute difference of 5.5 percentage points (95% CI 0.7–10.3). The direct-laryngoscopy success rate was substantially higher than in VODE. Differences in age, elective anaesthesia and operator experience help explain why the absolute improvement from one setting should not be transferred unchanged to another.⁴

Not every neonatal trial showed a clear advantage. Tippmann and colleagues' 2023 trial analysed 89 nasotracheal intubation encounters in 63 infants. First-attempt success was 20 of 41 with video laryngoscopy and 21 of 48 with direct laryngoscopy, giving an odds ratio of 1.22 (95% CI 0.51–2.88). The wide interval left substantial uncertainty. The different route and repeated encounters distinguish this small trial from VODE's oral intubations and once-only infant enrolment.⁵

The 2023 Cochrane review included eight studies contributing 759 neonatal intubation attempts. Video laryngoscopy increased first-attempt success in the pooled analysis (risk ratio 1.24; 95% CI 1.13–1.37), but certainty was low. Evidence for effects on desaturation or bradycardia was very uncertain. The review also combined different devices and clinical uses. VODE strengthens the evidence for procedural efficacy in urgent neonatal care, while the review's uncertainty about important physiological outcomes remains relevant.⁶

Oxygen delivery represents a separate opportunity to improve intubation. SHINE analysed 251 intubations in 202 infants at two Australian neonatal units. Nasal high-flow therapy increased successful first-attempt intubation without physiological instability from 31.5% to 50.0%, with an adjusted difference of 17.6 percentage points (95% CI 6.0–29.2). Its composite endpoint incorporated desaturation and bradycardia, whereas VODE's primary endpoint did not. These trials address complementary parts of airway management and do not establish the combined effect of their interventions.⁷

A further multicentre trial reported in 2023 compared direct and video laryngoscopy during elective intubation while both groups received supplemental oxygen. Among 244 analysed neonates and young infants, first-attempt success without desaturation was reported in 89.3% with video and 78.9% with direct laryngoscopy. This supports the possibility that a video advantage persists when oxygen is supplied, but the elective setting and different population limit extrapolation to VODE's urgent delivery-room and intensive care procedures.⁸

The range of control-group success rates also matters when applying these studies. A service achieving high success with direct laryngoscopy has less room for absolute improvement than one resembling VODE's control group. Equipment adoption should therefore be accompanied by assessment of local first-attempt success and complications, rather than an assumption that every unit will reproduce a 29-percentage-point improvement.¹

By the end of 2024, the evidence supported video laryngoscopy as an effective means of improving first-attempt success in neonatal practice, with VODE providing a particularly direct demonstration in urgent oral intubation. The remaining task is to translate that advantage into a consistently safe procedure: equipment access, training, supervision, oxygenation, timely interruption of unsuccessful attempts and confirmation of tube depth still matter. The trial supports incorporating video laryngoscopy into a prepared clinical system; it does not show that purchasing a screen alone eliminates the risks of neonatal intubation.

Should video laryngoscopy be used for urgent neonatal intubation?

VODE supports video laryngoscopy to improve first-attempt success in comparable neonatal settings, particularly where trainees perform intubation. A reduction in serious complications or long-term harm was not established.

References

1. 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.
2. VODE investigators. Supplementary appendix to: Video versus direct laryngoscopy for urgent intubation of newborn infants. *N Engl J Med* 2024;390(20):1885–1894.

3. O'Shea JE, Thio M, Kamlin CO, McGrory L, Wong C, John J, et al. Videolaryngoscopy to teach neonatal intubation: a randomized trial. *Pediatrics* 2015;136(5):912–919.
4. Garcia-Marcinkiewicz AG, Kovatsis PG, Hunyady AI, Olomu PN, Zhang B, Sathyamoorthy M, et al.; PeDI Collaborative investigators. First-attempt success rate of video laryngoscopy in small infants (VISI): a multicentre, randomised controlled trial. *Lancet* 2020;396(10266):1905–1913.
5. Tippmann S, Schäfer J, Winter J, Mühler AK, Schmitz K, Schönfeld M, et al. Video versus direct laryngoscopy to improve the success rate of nasotracheal intubations in the neonatal intensive care setting: a randomised controlled trial. *BMJ Paediatr Open* 2023;7(1):e001958.
6. Lingappan K, Neveln N, Arnold JL, Fernandes CJ, Pammi M. Videolaryngoscopy versus direct laryngoscopy for tracheal intubation in neonates. *Cochrane Database Syst Rev* 2023;5:CD009975.
7. 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.
8. Riva T, Engelhardt T, Basciani R, Bonfiglio R, Cools E, Fuchs A, et al.; OPTIMISE Collaboration. Direct versus video laryngoscopy with standard blades for neonatal and infant tracheal intubation with supplemental oxygen: a multicentre, non-inferiority, randomised controlled trial. *Lancet Child Adolesc Health* 2023;7(2):101–111.

PLUSS

Manley BJ, Kamlin COF, Donath SM, Francis KL, Cheong JLY, Dargaville PA, et al.; PLUSS Trial Investigators. Intratracheal budesonide mixed with surfactant for extremely preterm infants: the PLUSS randomized clinical trial. JAMA 2024;332(22):1889–1899.

Introduction

Survival after extremely preterm birth can be accompanied by prolonged dependence on respiratory support. Bronchopulmonary dysplasia (BPD) reflects injury and disrupted development in an immature lung, with consequences extending beyond the neonatal admission. Preventing it requires more than correction of the initial surfactant deficiency: inflammation, oxygen exposure and mechanical ventilation contribute to a process that develops over weeks. A treatment delivered during the first hours of respiratory distress could conceivably modify that trajectory, but an early improvement in pulmonary physiology would not by itself establish a lasting clinical benefit.¹

Corticosteroids offer a plausible anti-inflammatory intervention, while their potential effects on other developing organs make the route, dose and timing especially important. Mixing budesonide with surfactant provides a means of delivering the drug directly into the lung during an intervention already indicated for respiratory distress. This approach is distinct from prolonged inhaled treatment and from systemic corticosteroid courses. It nevertheless requires evidence of both effectiveness and safety; intratracheal delivery cannot be assumed to eliminate systemic exposure or delayed adverse effects.¹

Earlier trials reported substantial reductions in death or chronic lung disease with this combination. Those results created an important opportunity for replication in a larger, more diverse population receiving contemporary neonatal care. The Preventing Lung Disease Using Surfactant + Steroid (PLUSS) trial enrolled infants born before 28 weeks who required surfactant within the first 48 hours. It tested whether adding budesonide improved survival without physiological BPD at 36 weeks' postmenstrual age. The central question was whether the intervention produced a clinically meaningful benefit across this population, with particular care needed when interpreting smaller subgroups and outcomes conditional on survival.¹

Synopsis

PLUSS was a double-blind randomised trial conducted in 21 tertiary neonatal units in Australia, New Zealand, Canada and Singapore. Recruitment ran from January 2018 to March 2023. Eligible infants were born before 28 weeks' gestation, were younger than 48 hours, and either received invasive ventilation or received non-invasive respiratory support with a clinical decision to administer surfactant. Invasively ventilated infants did not need a specified oxygen requirement. One previous surfactant dose was permitted, with eligibility for further treatment assessed six to 12 hours later. Important exclusions included previous postnatal corticosteroid treatment to prevent lung disease, major anomalies affecting respiration, and circumstances in which survival or continuing active care was considered unlikely.¹

Of 2,836 infants screened, 2,276 were eligible and 1,062 were randomised: 525 to budesonide plus surfactant and 537 to surfactant alone. Parents subsequently withdrew permission to retain data for one and two infants, respectively, leaving 524 and 535 in the analysis. Other infants who did not receive their assigned intervention remained in their allocated groups. Allocation was concealed and stratified by centre, gestational-age category, previous surfactant treatment and respiratory support at randomisation. Infants from multiple pregnancies were randomised individually. Two staff members outside the infant's care team prepared the intervention privately and covered the syringe, maintaining masking of treating clinicians, parents and outcome assessors.¹

The active treatment contained budesonide 0.25 mg/kg mixed with poractant alfa. The first study dose contained surfactant 200 mg/kg; a second, containing surfactant 100 mg/kg and the same budesonide dose, was given six to 12 hours later if eligibility criteria persisted. The comparator used identical surfactant doses without budesonide. Administration could occur through an endotracheal tube during ongoing ventilation, during an intubate–surfactant–extubate procedure, or through a thin catheter during non-invasive support. The second study dose could be administered after 48 hours of age. Inhaled or nebulised corticosteroids were prohibited before assessment of the primary outcome, but subsequent systemic corticosteroids and other care followed local practice.¹

The primary outcome was survival without physiological BPD, assessed between 36 weeks and 36 weeks six days' postmenstrual age. BPD included invasive ventilation, non-invasive positive-pressure support or nasal high flow of at least 2 L/min, regardless of oxygen concentration. Infants receiving lower-flow oxygen met the definition if effective FiO_2 was at least 0.30, or if they failed an air-reduction assessment at a lower effective concentration. Infants discharged home earlier without oxygen or respiratory support were classified as free of BPD. When assessment was incomplete or incorrectly

timed, masked reviewers applied a separate algorithm to information recorded at exactly 36 weeks. This was required for 102 of 867 survivors.^{1,2}

The target sample provided 90% power to detect an absolute increase in survival without BPD from 50% to 60%, with allowance for losses. The principal analysis estimated an adjusted risk difference, accounting for gestational-age category, previous surfactant and respiratory support at randomisation. Standard errors allowed for infants from the same pregnancy. Centre was used in randomisation but was not included among these three adjustment variables. Secondary analyses evaluated survival, BPD among survivors, respiratory support, other neonatal morbidities, growth and adverse events. Safety monitoring and an interim efficacy assessment did not lead to early termination.¹

Mean gestational age was 25.6 weeks and mean birth weight 775 g. Approximately 55% of infants were male, three quarters received invasive ventilation at randomisation, and 57% had already received surfactant. Almost all had some antenatal corticosteroid exposure. Median age at randomisation was 4.6 hours with budesonide and 5.0 hours with control. Median FiO₂ immediately before intervention was 0.30 in both groups. Most study doses were administered through an endotracheal tube during ongoing ventilation; approximately 14% used a thin catheter. A second study dose was given to 68.5% and 72.7%, respectively.¹

Survival without BPD occurred in 134 of 524 infants receiving budesonide and 121 of 535 controls, or 25.6% versus 22.6%. The adjusted risk difference was 2.7 percentage points (95% CI -2.1 to 7.4). Survival at 36 weeks was 83.2% versus 80.6%, with a reported adjusted difference of 1.4 percentage points (95% CI -2.9 to 5.7). Among survivors, BPD occurred in 302 of 436 versus 310 of 431, or 69.3% versus 71.9%; the reported difference was -2.7 percentage points (95% CI -8.4 to 3.1). None of these comparisons established benefit.¹

Median duration of invasive ventilation was eight days in both groups, and median hospital stay among survivors was 110 days in both. Subsequent systemic corticosteroids were used in 33.2% and 32.9%. There was no convincing reduction in other reported respiratory outcomes. Hyperglycaemia or insulin treatment within 14 days occurred in 64.9% versus 61.9%, a difference of 3.0 percentage points (95% CI -2.7 to 8.8). Spontaneous intestinal perforation occurred in 19 versus 17 infants, a difference of 0.4 percentage points (95% CI -1.7 to 2.6). The safety findings did not demonstrate a clear excess of these events, while leaving clinically relevant uncertainty.^{1,2}

A post hoc analysis examined oxygen requirements immediately before the first intervention. Among 226 infants receiving FiO_2 of at least 0.50, survival without BPD occurred in 21 of 104 versus 15 of 122, with an adjusted difference of 10.2 percentage points (95% CI 1.5 to 18.8). The corresponding adjusted difference in infants receiving lower FiO_2 was 0.4 percentage points (95% CI -5.1 to 5.8). The interaction P value was 0.07. The overall trial therefore did not confirm improved survival without BPD, and the oxygen-defined subgroup finding remained exploratory.¹

Critique

PLUSS provides substantially more reliable evidence for this treatment strategy than a small efficacy study could provide. Its multicentre design, concealed allocation and practical masking procedures reduce important sources of bias. Masking is especially valuable when respiratory support, additional corticosteroids and readiness for discharge depend partly on clinical judgement. Only three randomised infants were removed because permission to use their data was withdrawn, and non-receipt of assigned treatment did not otherwise redefine the comparison. These features support an interpretation based on allocation rather than on selecting infants who tolerated or completed treatment.¹

The primary result needs to be read through its confidence interval. It does not establish equivalence, nor does it exclude every worthwhile benefit. The interval permits a modest improvement in survival without BPD and a smaller absolute worsening. It does, however, exclude the ten-percentage-point improvement used to design the study. Describing PLUSS simply as underpowered because the control success rate was lower than expected is misleading: a lower probability of this favourable binary outcome does not necessarily reduce precision for a fixed absolute difference. A proportionate improvement would become a smaller absolute effect at this baseline risk, and that is a different claim.¹

Combining survival and BPD is clinically appropriate because an infant who dies cannot subsequently be assessed for chronic lung disease. Reporting BPD only among survivors would risk presenting a favourable pulmonary outcome if more severely affected infants died. PLUSS reports both components, and neither established an independent benefit. Equally, the numerically greater survival in the budesonide group is insufficient to support a mortality claim. Its uncertainty must remain visible alongside the uncertainty around BPD. These data support a neutral overall conclusion without converting either component into a separate positive trial.¹

The outcome definition also carries practical limitations. Respiratory support at 36 weeks combines underlying disease with local treatment and weaning practices. Physiological testing helps standardise lower oxygen requirements, but about one in eight survivors needed committee adjudication because the planned assessment was absent, incomplete or mistimed. The replacement algorithm used a lower effective oxygen threshold of 0.22 when classifying infants on low-flow or ambient oxygen, rather than reproducing an air-reduction challenge. Masked review reduces the risk of differential classification, but does not make the two assessment pathways identical. This introduces measurement uncertainty without providing evidence that it favoured either group.^{1,2}

The enrolled population was broader than a population with severe early respiratory distress requiring high oxygen concentrations. Median FiO_2 was only 0.30, non-invasive support was permitted, and prior surfactant was common. These characteristics improve relevance to many contemporary neonatal units while limiting direct comparison with earlier, more selective trials. They also mean that a hypothesis about benefit in severe inflammatory lung disease was not tested exclusively in that group. Conversely, the absence of a severity threshold should not be used after the result to dismiss a well-conducted trial of the actual clinical strategy under consideration.¹

The high- FiO_2 finding is a plausible hypothesis, but an inadequate basis for routine targeted treatment. The analysis was post hoc, the interaction did not provide conventional statistical evidence of different treatment effects, and only 36 infants in that subgroup survived without BPD. A significant confidence interval within one subgroup and a non-significant interval within another do not establish that the subgroups respond differently. Furthermore, oxygen requirements depend on respiratory support and treatment decisions as well as lung disease. The measurement immediately before intervention is not a validated biological selection test. Confirmation would require a prospective comparison capable of testing the proposed treatment interaction.¹

Timing provides another possible explanation that remains unproven. More than half the infants had received a previous surfactant dose, and the first study intervention could therefore represent repeat treatment. Yet the trial also included 458 infants without previous surfactant, who were randomised much earlier, at a median age of approximately one hour. There was no clear interaction according to prior surfactant exposure. That subgroup analysis cannot establish that every early-treatment strategy is ineffective, but it weakens a simple claim that the overall result is explained entirely by delayed administration. Comparisons between infants with and without previous surfactant are additionally confounded by their different illness severity and maturity.^{1,2}

Subsequent systemic corticosteroid use in roughly one third of both groups needs a similarly careful interpretation. This was permitted care, so the trial estimated the effect of adding early intratracheal budesonide to a pathway that could include later systemic treatment. Such treatment might reduce longer-term separation between groups, but withholding it would have tested a different clinical strategy. The absence of a reduction in subsequent systemic corticosteroid use also means that PLUSS did not demonstrate a steroid-sparing advantage. Delivery through different airway techniques adds practical breadth, although the smaller thin-catheter subgroup cannot establish effectiveness for that method independently.¹

Safety remains an unfinished part of the clinical question at the book's evidence cutoff. The trial did not identify a convincing increase in its reported neonatal adverse events, but several were uncommon and their confidence intervals permit meaningful harm. A high background frequency of hyperglycaemia also makes a modest drug-associated increase difficult to exclude. Normality of short-term growth or cranial ultrasound outcomes, where assessed, would not substitute for later neurodevelopmental assessment. Two-year follow-up was not available in this report. The absence of an immediate safety signal therefore supports continued evaluation without establishing that the intervention is harmless.^{1,2}

Finally, fewer than half of eligible infants were randomised. Parental decisions, unavailable staff and missed opportunities all contributed. This does not undo randomisation within the recruited cohort, but it qualifies claims that the results describe every extremely preterm infant who might receive surfactant. Units, surfactant formulations and treatment thresholds may also differ elsewhere. The principal conclusion is nonetheless directly useful: in the studied international population, adding one or two doses of budesonide to poractant alfa did not deliver the large improvement anticipated. Selective adoption based on an exploratory subgroup would exceed what these data establish.¹

Body of Evidence

The 2008 pilot study by Yeh and colleagues randomised 116 infants weighing less than 1,500 g who required mechanical ventilation and FiO_2 of at least 0.60 shortly after birth. Budesonide 0.25 mg/kg was combined with surfactant, with repeat doses permitted. Death or chronic lung disease occurred in 19 of 60 intervention infants and 34 of 56 controls. This substantial difference, together with improved early respiratory measures, justified further investigation. The small sample and selected population, however, could not establish broad effectiveness or exclude uncommon and delayed adverse effects.³

The subsequent 2016 trial enrolled 265 very-low-birth-weight infants in three centres in Taiwan and the United States. Eligibility required severe respiratory distress, invasive ventilation and FiO_2 of at least 0.50 within four hours of birth. Death or BPD occurred in 55 of 131 infants receiving budesonide plus surfactant and 89 of 134 receiving surfactant alone. The direction and magnitude reinforced the earlier findings, but the population, surfactant preparation and repeat-treatment approach differed from PLUSS. Those differences support investigation of effect modifiers; they do not establish which feature caused the discrepancy between trials.⁴

NEUROSIS examined a different delivery strategy. In 2015, 863 extremely preterm infants were randomised to early inhaled budesonide or placebo, with treatment continuing until respiratory support was no longer required or 32 weeks' postmenstrual age. Death or BPD occurred in 40.0% versus 46.3% of infants with adequate primary-outcome data, with a stratified relative risk of 0.86 (95% CI 0.75 to 1.00). BPD was reduced, but mortality was numerically higher. This divergence illustrated why a pulmonary benefit must be interpreted together with survival rather than being treated as sufficient evidence for adoption.⁵

Follow-up of NEUROSIS reported no established difference in neurodevelopmental disability among assessed survivors at 18 to 22 months. Mortality among infants with known vital status was 82 of 413 versus 58 of 400, or 19.9% versus 14.5% (relative risk 1.37; 95% CI 1.01 to 1.86). This later finding reinforces the need for follow-up that includes deaths as well as outcomes among survivors. It does not prove that the much shorter intratracheal regimen used in PLUSS causes the same harm: route and cumulative exposure were different.⁶

Systemic corticosteroid evidence also resists a single class-wide conclusion. PREMILOC randomised 523 extremely preterm infants to ten days of low-dose hydrocortisone or placebo, analysing 521 after consent withdrawals. Survival without BPD occurred in 60% versus 51%, with an adjusted odds ratio of 1.48 (95% CI 1.02 to 2.16). The study investigated a specific early systemic regimen and reported a sepsis signal in its least mature subgroup. Its findings cannot be transferred directly to intratracheal budesonide, but show why the compound, exposure and population matter when comparing neonatal steroid trials.⁷

The contemporaneous editorial accompanying PLUSS emphasised the importance of large confirmatory neonatal trials and the uncertainty surrounding explanations based on disease severity or treatment timing. It also identified the Budesonide in Babies trial as a source of forthcoming evidence. At the end of 2024, its results were not available for this assessment. The appropriate evidence synthesis therefore places PLUSS

alongside earlier trials without importing subsequent findings or treating an anticipated individual-participant meta-analysis as if it had already resolved the subgroup question.⁸

Taken together, the evidence supports a narrower conclusion than either universal enthusiasm or universal rejection of neonatal corticosteroids. Earlier intratracheal studies generated a strong efficacy signal, while PLUSS did not reproduce a large benefit in a broader population. Inhaled and systemic regimens answer related but distinct questions. Further evidence needs to address reproducibility, prospectively defined treatment selection and survival with acceptable long-term development. Until then, the proposed advantages of adding budesonide routinely to surfactant remain unconfirmed.¹

Should budesonide routinely be added to surfactant for extremely preterm infants?

PLUS does not support routine addition: it did not establish improved survival without BPD. Modest benefit remains possible, but the high-oxygen subgroup finding and long-term safety require further evidence.

References

1. Manley BJ, Kamlin COF, Donath SM, Francis KL, Cheong JLY, Dargaville PA, et al.; PLUS Trial Investigators. Intratracheal budesonide mixed with surfactant for extremely preterm infants: the PLUS randomized clinical trial. *JAMA* 2024;332(22):1889–1899.
2. PLUS Trial Investigators. Supplement 2 to: Intratracheal budesonide mixed with surfactant for extremely preterm infants: the PLUS randomized clinical trial. *JAMA* 2024;332(22):1889–1899.
3. Yeh TF, Lin HC, Chang CH, Wu TS, Su BH, Li TC, et al. Early intratracheal instillation of budesonide using surfactant as a vehicle to prevent chronic lung disease in preterm infants: a pilot study. *Pediatrics* 2008;121(5):e1310–e1318.
4. Yeh TF, Chen CM, Wu SY, Husan Z, Li TC, Hsieh WS, et al. Intratracheal administration of budesonide/surfactant to prevent bronchopulmonary dysplasia. *Am J Respir Crit Care Med* 2016;193(1):86–95.
5. Bassler D, Plavka R, Shinwell ES, Hallman M, Jarreau PH, Carnielli V, et al.; NEUROSIS Trial Group. Early inhaled budesonide for the prevention of bronchopulmonary dysplasia. *N Engl J Med* 2015;373(16):1497–1506.

6. Bassler D, Shinwell ES, Hallman M, Jarreau PH, Plavka R, Carnielli V, et al.; Neonatal European Study of Inhaled Steroids Trial Group. Long-term effects of inhaled budesonide for bronchopulmonary dysplasia. *N Engl J Med* 2018;378(2):148–157.
7. Baud O, Maury L, Lebail F, Ramful D, El Moussawi F, Nicaise C, et al.; PREMILOC trial study group. Effect of early low-dose hydrocortisone on survival without bronchopulmonary dysplasia in extremely preterm infants (PREMILOC): a double-blind, placebo-controlled, multicentre, randomised trial. *Lancet* 2016;387(10030):1827–1836.
8. Jensen EA. Intratracheal budesonide combined with surfactant in extremely preterm infants. *JAMA* 2024;332(22):1885–1887.

Possible Reporting Inconsistencies. The text gives 418 infants in the lower-FiO₂ subgroup, whereas Figure 2 shows 411 plus 407, totalling 818. In Table 2, the control-group respiratory-support rows print denominators of 436; their counts sum to 431 and their percentages correspond to 431 survivors. These discrepancies do not alter the reported primary-outcome counts.



Perioperative Trials

OPTIMISE II

OPTIMISE II Trial Group. Cardiac output-guided haemodynamic therapy for patients undergoing major gastrointestinal surgery: OPTIMISE II randomised clinical trial. BMJ 2024;387:e080439.

Introduction

Haemodynamic management during major surgery aims to maintain organ perfusion without causing harm through excess fluid or unnecessary cardiovascular stimulation. Blood pressure, heart rate and urine output provide useful information, but do not directly describe cardiac output or the response to a fluid challenge. Monitoring stroke volume therefore offers an attractive way to individualise fluid administration. The clinical value of that information depends on the decisions it triggers and on whether those decisions improve outcomes beyond contemporary usual care.¹

Older patients undergoing major gastrointestinal surgery have a substantial risk of infection and other postoperative complications. Improving tissue perfusion might support wound healing, preserve intestinal barrier function and reduce infection. Earlier studies of goal-directed therapy combined physiological targets with fluid administration and, in some cases, inotropes. Their apparent benefits were difficult to interpret across differing algorithms, monitoring devices, populations and outcome definitions. Developments in surgery and perioperative care also changed the comparator against which an additional monitoring strategy needed to demonstrate value.¹

OPTIMISE II tested a specific treatment package: minimally invasive cardiac-output monitoring, stroke-volume-guided fluid boluses and a fixed low-dose inotrope infusion, continued during surgery and for four hours afterwards. Its principal aim was to prevent postoperative infection, while prospectively collected cardiac outcomes addressed the possibility that physiological optimisation could itself cause harm. The trial was larger and more international than its predecessor, with nearly 2,500 participants across 55 hospitals. Its interpretation turns on two distinctions: failure of this package is not a test of every possible use of cardiac-output monitoring, and a neutral infection result does not make an early cardiac safety signal clinically irrelevant.¹

Synopsis

OPTIMISE II was a randomised, observer-masked effectiveness trial conducted in 11 countries between January 2017 and September 2022. Patients were at least 65 years old, had an American Society of Anesthesiologists (ASA) physical status of II or higher, and were scheduled for elective major gastrointestinal surgery expected to last more than 90 minutes. Eligible procedures involved gut resection, including colorectal, gastric, oesophageal and pancreatic operations. Important exclusions included recent acute myocardial ischaemia or pulmonary oedema, contraindications to low-dose inotropes, and a life expectancy of less than 30 days. Emergency surgery was outside the trial's population.¹

Of 6,693 patients assessed, 2,502 were randomised. Four participants were subsequently excluded because they did not meet the age criterion, following an independent recommendation made without knowledge of allocation. The resulting groups comprised 1,251 intervention patients and 1,247 controls. Randomisation used a central online service and minimisation with a random component, balancing country, surgical procedure and ASA category. Clinicians administering treatment could not be masked. Outcome collectors were separate from the care team and intended to remain unaware of allocation, with clinical outcomes confirmed by a masked local investigator or delegate.¹

Both groups received general perioperative measures intended to avoid extremes of practice, including targets for oxygenation, haemoglobin, temperature, heart rate and arterial pressure. Maintenance crystalloid was recommended at 1 mL/kg/h, while clinicians selected fluids for volume expansion and provided anaesthesia, analgesia and antibiotic prophylaxis according to local practice. The intervention began after induction of anaesthesia. Stroke volume was measured using an Edwards EV1000 or HemoSphere system with either a FloTrac arterial waveform sensor or a ClearSight finger cuff. Clinicians chose the sensor rather than being randomised between technologies.¹

Additional 250 mL fluid boluses were guided by the stroke-volume response. A rise of at least 10% indicated continuing responsiveness; further boluses were withheld when stroke volume no longer increased and remained stable, or when stroke-volume variation indicated a low likelihood of responsiveness. After the first bolus, patients received a fixed infusion of dobutamine 2.5 µg/kg/min or dopexamine 0.5 µg/kg/min. Sustained tachycardia despite adequate anaesthesia and analgesia prompted dose reduction and, if persistent, termination. Although either drug was permitted, dobutamine was the agent actually used. The algorithm continued for four postoperative hours.¹

Controls received usual haemodynamic care without routine cardiac-output monitoring. Clinicians could give fluid boluses using conventional clinical and laboratory information and could administer vasoactive drugs when indicated. Cardiac-output monitoring could be requested if a control patient became critically unwell. The comparison therefore tested addition of a specified algorithm and routine inotrope, rather than comparing active cardiovascular management with no treatment. Adherence was assessed from recorded delivery of monitoring and inotrope treatment, and from any use of cardiac-output monitoring in controls.^{1,2}

The primary outcome was at least one postoperative infection within 30 days of randomisation, using Centers for Disease Control and Prevention definitions. Components included superficial, deep or organ-space surgical-site infection, pneumonia, urinary tract infection, bloodstream infection and infection of uncertain source. Events had to require clinical treatment, corresponding to Clavien–Dindo grade II or higher. Safety outcomes were acute cardiac events within 24 hours and within 30 days, incorporating arrhythmia, myocardial infarction, myocardial injury, successfully resuscitated cardiac arrest and cardiogenic pulmonary oedema. Secondary outcomes included acute kidney injury of KDIGO stage 2 or greater within 30 days and mortality within 180 days.¹

The planned sample of 2,502 provided more than 90% power to detect a reduction in infection from 30% to 24%. The principal analysis used mixed-effects logistic regression, with a country random intercept and adjustment for procedure, age, sex, ASA category, haemoglobin and creatinine. Patients with observed outcomes were analysed according to allocation. Additional analyses examined death as a competing risk and a simpler adjustment model. Mean age was 74 years, 57% were women, and approximately half had ASA III status. About 84% were scheduled for cancer surgery. Colorectal or small-bowel resection predominated, and approximately half underwent laparoscopic or laparoscopic-assisted surgery.¹

Primary-outcome information was available for 1,247 patients in each group. Infection occurred in 289 intervention patients and 283 controls, or 23.2% versus 22.7% (adjusted odds ratio 1.03; 95% CI 0.84 to 1.25; $P=0.81$). The adjusted absolute difference was 0.4 percentage points (95% CI –3.4 to 4.2). The competing-risk and minimally adjusted analyses did not materially alter the conclusion. There was no clear treatment interaction by planned procedure, and post hoc surgical-approach analyses did not identify a convincing differential effect.^{1,2}

Acute cardiac events within 24 hours occurred in 38 of 1,250 intervention patients and 21 of 1,247 controls, or 3.0% versus 1.7% (adjusted odds ratio 1.82; 95% CI 1.06 to 3.13;

P=0.03). Arrhythmias accounted for most of the difference, with 33 versus 17 affected patients. At 30 days, the broader cardiac outcome occurred in 6.8% versus 6.3% (adjusted odds ratio 1.06; 95% CI 0.77 to 1.47). Acute kidney injury occurred in 40 versus 32 patients. Mortality by 180 days was 68 of 1,241 versus 84 of 1,239, with an adjusted odds ratio of 0.76 (95% CI 0.54 to 1.07). Neither secondary outcome established benefit. Hospital stay among survivors and days alive outside critical care were also similar.^{1,2}

Critique

The trial's size, international recruitment and nearly complete primary-outcome ascertainment are major strengths. The central allocation process balanced important prognostic factors, and the analysis appropriately retained patients in their assigned groups despite treatment deviations. Excluding four ineligible younger participants means that the analysis was not literally inclusive of everyone randomised, but the small number and allocation-blind independent decision make substantial bias from this exclusion unlikely. Missing primary-outcome data were limited to four intervention patients. These features make large effects concealed by attrition or extensive crossover an implausible explanation for the neutral finding.¹

The primary confidence interval is more informative than the non-significant P value alone. On the absolute scale it excludes the six-percentage-point reduction used for sample-size planning, while allowing smaller benefit or harm. The study was a superiority trial and did not establish equivalence. Its control infection rate was below the anticipated 30%, but this observation does not automatically establish inadequate power for the original absolute target. It does indicate that the expected improvement was ambitious relative to the risk actually observed. No clinically important benefit was demonstrated to offset the additional treatment and monitoring burden.¹

Intervention delivery provides evidence that a meaningful treatment contrast occurred. Dobutamine was administered to 94% of intervention patients and only six controls. Recorded deviations affected 7.4% versus 1.4%, including some overlapping failures of monitoring and inotrope delivery. The fluid contrast was much smaller: reported total volumes across the intervention period were approximately 3.27 L versus 3.06 L, and balanced crystalloids predominated. Similar aggregate volumes do not mean that timing or patient-level allocation was identical, but they limit any claim that the intervention worked by producing a large change in total fluid exposure. Detailed physiological responses were not reported.^{1,2}

This distinction also limits mechanistic attribution. The trial randomised a combined monitoring-and-treatment strategy. It did not separately randomise monitoring, fluid

optimisation or dobutamine, and therefore cannot estimate the independent effect of each component. Dobutamine is a plausible contributor to the excess early arrhythmias, particularly given the marked exposure difference and frequent dose reductions for tachycardia. That remains an interpretation of the package's effects, not a factorial demonstration of causation. Equally, the neutral infection outcome cannot establish that cardiac-output information is useless when investigating instability or guiding a different, selectively applied treatment strategy.¹

The safety result warrants attention without exaggeration. The absolute excess in the 24-hour cardiac composite was 1.3 percentage points (95% CI 0.1 to 2.5). Most arrhythmias were managed pharmacologically, but the severity tables also included life-threatening and fatal classifications. They should not all be dismissed as transient monitor abnormalities. Conversely, the comparison was one of several reported outcomes and its P value was nominal, without a stated multiplicity correction. The combination of biological plausibility, temporal proximity and absence of demonstrated infection benefit makes the signal relevant to routine adoption, while leaving uncertainty about its precise size.^{1,2}

The later cardiac comparison does not erase events occurring soon after treatment. Both groups accumulated further events, and the 30-day confidence interval was compatible with benefit and harm. In addition, myocardial injury after non-cardiac surgery was not sought through routine screening of every participant. Its true incidence may therefore have been greater than the clinically detected count. This limits reassurance about subclinical injury and prevents treating the safety composite as a complete account of cardiovascular effects. The numerically lower 180-day mortality in the intervention group is likewise insufficient to establish a survival advantage or to neutralise the early arrhythmia finding.¹

Masking was incomplete in practice. Assessors considered themselves suitably blinded for approximately 70% of evaluations; around 18% reported knowing allocation and another 11% might have known it. Standard definitions and independent confirmation help, but an infection endpoint requiring treatment still depends partly on investigations and decisions by unmasked clinicians. The same consideration applies to arrhythmias requiring treatment. Antibiotic prophylaxis followed local guidance, and adherence to those policies was not recorded. These limitations introduce uncertainty about ascertainment and co-interventions, although the available data do not identify a systematic imbalance sufficient to explain the principal result.^{1,2}

The sensor analyses require particular restraint. FloTrac was used far more often than ClearSight, and clinicians selected the sensor after randomisation. The supplement

compared each intervention sensor subgroup with the entire control group, explicitly describing these as pseudo-subgroup analyses. Thus, the lower infection rate among patients selected for ClearSight is not a randomised demonstration that it is superior to FloTrac or beneficial compared with equivalent controls. Patients selected for a less invasive device may have differed in baseline risk or procedure. Concerns about monitoring precision are reasonable hypotheses, but this trial cannot determine whether another sensor would have changed its outcome.^{1,2}

Generalisability also depends on the actual population and delivery resources. Participants were older elective surgical patients, predominantly undergoing cancer-related colorectal or small-bowel procedures. Emergency laparotomy, recent unstable cardiac disease and operations outside the defined gastrointestinal categories were not represented. Research staff assisted with intervention delivery in about half of operations and two fifths of postoperative periods. That support may exceed what is routinely available and did not yield an established clinical gain. At the same time, the extensive use of minimally invasive surgery and contemporary fluid practice makes the comparison more relevant to present-day elective care than many older efficacy studies.¹

Industry involvement was disclosed: Edwards Lifesciences and the UK National Institute for Health and Care Research funded the trial, and several investigators reported relationships with the manufacturer. The report states that funders had no role in design, data collection, analysis, interpretation or writing. These disclosures merit transparency, while the neutral effectiveness result and reported safety concern should be judged on the methods and data. Overall, OPTIMISE II argues against routine application of this fixed-inotrope package to broadly selected older gastrointestinal surgical patients. It leaves open more focused questions about treating demonstrated haemodynamic abnormalities.¹

Body of Evidence

The original OPTIMISE trial enrolled 734 higher-risk patients undergoing major gastrointestinal surgery in 17 UK hospitals. Cardiac-output-guided fluid therapy with dopexamine continued during surgery and for six hours afterwards. The primary composite of moderate or major complications and mortality occurred in 36.6% versus 43.4% (relative risk 0.84; 95% CI 0.71 to 1.01; P=0.07). Five intervention patients experienced cardiovascular serious adverse events within 24 hours, compared with none receiving usual care. The trial generated a credible hypothesis, but did not itself establish superiority for its primary outcome.³

POEMAS tested a non-invasive cardiac-output-guided strategy in 142 patients undergoing major abdominal surgery across six hospitals. The intervention included fluid and vasoactive treatment rather than a monitoring device in isolation. Overall complications occurred in approximately 40% versus 41%, with no established improvement in hospital stay. Its modest sample left considerable uncertainty, but the neutral primary findings were a reminder that favourable results from selected small studies were not consistently reproduced in pragmatic settings. This was a different algorithm and should not be treated as an exact replication of OPTIMISE II.⁴

In a 2021 Dutch multicentre trial, de Waal and colleagues studied 482 patients having high-risk elective abdominal surgery. An age-specific algorithm used cardiac index and stroke-volume variation during a 24-hour perioperative period. The mean number of major complications per patient was 0.79 with goal-directed treatment and 0.69 with control ($P=0.195$). Intervention patients received more intraoperative fluid and more dobutamine, without established improvements in complications or length of stay. This trial also challenged the assumption that achieving specified haemodynamic targets necessarily translates into better clinical outcomes.⁵

A 2022 systematic review by Jessen and colleagues included a main pool of 76 trials. It estimated a mortality odds ratio of 0.84 (95% CI 0.64 to 1.09) and a reduction in hospital stay of 0.72 days, both supported by low-certainty evidence. Infectious complications and anastomotic leakage were among the outcomes with moderate-certainty evidence. The review explains why another definitive trial remained justified, while its varied interventions and generally intermediate risk of bias limit direct application of pooled findings to one specific package. It preceded OPTIMISE II and cannot be treated as an updated synthesis incorporating it.⁶

RELIEF addressed fluid quantity rather than cardiac-output-guided optimisation. Among 3,000 patients assigned to restrictive or liberal perioperative fluid regimens, disability-free survival at one year was similar, while acute kidney injury occurred in 8.6% versus 5.0%. This result does not favour the OPTIMISE II algorithm, but cautions against translating its neutral findings into indiscriminate fluid restriction. Both fluid excess and inadequate replacement remain clinically relevant possibilities, and a trial of one haemodynamic package does not determine the appropriate volume for every patient.⁷

INPRESS examined individualised blood-pressure management in a smaller, higher-risk surgical population. Of 298 randomised patients, 292 were analysed. Targeting systolic pressure close to the patient's resting value reduced the primary composite of systemic inflammatory response and organ dysfunction by day seven: 38.1% versus 51.7% (relative risk 0.73; 95% CI 0.56 to 0.94). Its strategy, comparator and endpoint differed

from OPTIMISE II. The finding supports further evaluation of personalised cardiovascular treatment, without establishing that routine cardiac-output optimisation plus an inotrope is effective.⁸

The accompanying 2024 editorial considered improvements in usual perioperative care, possible monitoring limitations and dobutamine-associated arrhythmias as explanations for OPTIMISE II. These are interpretations rather than separately randomised comparisons. The evidence available at the end of 2024 therefore supports a specific practice conclusion: routine use of this package lacks demonstrated benefit and carries an early safety concern. It does not resolve targeted monitoring in unstable patients, alternative pressure-based strategies, or emergency surgery. Those questions require evidence matched to their own populations and interventions.⁹

Should this cardiac-output-guided fluid and fixed-inotrope algorithm be used routinely for older patients undergoing elective gastrointestinal surgery?

No. OPTIMISE II found no reduction in postoperative infection and reported more early cardiac events, chiefly arrhythmias; it does not determine the value of selectively monitoring or treating specific haemodynamic abnormalities.

References

1. OPTIMISE II Trial Group. Cardiac output-guided haemodynamic therapy for patients undergoing major gastrointestinal surgery: OPTIMISE II randomised clinical trial. *BMJ* 2024;387:e080439.
2. OPTIMISE II Trial Group. Supplementary appendix to: Cardiac output-guided haemodynamic therapy for patients undergoing major gastrointestinal surgery: OPTIMISE II randomised clinical trial. *BMJ* 2024;387:e080439.
3. Pearse RM, Harrison DA, MacDonald N, Gillies MA, Blunt M, Ackland G, et al.; OPTIMISE Study Group. Effect of a perioperative, cardiac output-guided hemodynamic therapy algorithm on outcomes following major gastrointestinal surgery: a randomized clinical trial and systematic review. *JAMA* 2014;311(21):2181–2190.
4. Pestaña D, Espinosa E, Eden A, Nájera D, Collar L, Aldecoa C, et al. Perioperative goal-directed hemodynamic optimization using noninvasive cardiac output monitoring in major abdominal surgery: a prospective, randomized, multicenter, pragmatic trial: POEMAS Study (PeriOperative goal-directed thErapy in Major Abdominal Surgery). *Anesth Analg* 2014;119(3):579–587.

5. de Waal EEC, Frank M, Scheeren TWL, Kaufmann T, de Korte-de Boer D, Cox B, et al. Perioperative goal-directed therapy in high-risk abdominal surgery. A multicenter randomized controlled superiority trial. *J Clin Anesth* 2021;75:110506.
6. Jessen MK, Vallentin MF, Holmberg MJ, Bolther M, Hansen FB, Holst JM, et al. Goal-directed haemodynamic therapy during general anaesthesia for noncardiac surgery: a systematic review and meta-analysis. *Br J Anaesth* 2022;128(3):416–433.
7. Myles PS, Bellomo R, Corcoran T, Forbes A, Peyton P, Story D, et al.; Australian and New Zealand College of Anaesthetists Clinical Trials Network and the Australian and New Zealand Intensive Care Society Clinical Trials Group. Restrictive versus liberal fluid therapy for major abdominal surgery. *N Engl J Med* 2018;378(24):2263–2274.
8. Futier E, Lefrant JY, Guinot PG, Godet T, Lorne E, Cuvillon P, et al.; INPRESS Study Group. 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.
9. Fischer MO, Lorne E. Haemodynamic management during major surgery. *BMJ* 2024;387:q2593.

TIGHT K

O'Brien B, Campbell NG, Allen E, Jamal Z, Sturgess J, Sanders J, et al.; TIGHT K investigators. Potassium supplementation and prevention of atrial fibrillation after cardiac surgery: the TIGHT K randomized clinical trial. JAMA 2024;332(12):979–988.

Introduction

Potassium supplementation is a familiar part of postoperative cardiac care. Correcting clinically important hypokalaemia has a clear physiological rationale, but maintaining potassium in the upper part of the normal range is a separate intervention. It requires repeated decisions, drug preparation and administration, often through intravenous access. A common practice can therefore consume substantial clinical time even when each individual dose seems inexpensive and its expected benefit is rarely questioned.¹

Atrial fibrillation after cardiac surgery is frequent and associated with longer admissions and adverse outcomes. This has encouraged preventive strategies directed at several possible mechanisms, including electrolyte concentrations. Many units supplement potassium whenever the serum concentration falls below 4.5 mEq/L, aiming to prevent postoperative atrial arrhythmias. The inference that avoiding a low concentration must justify actively maintaining a high-normal concentration, however, requires testing. The two policies differ in treatment burden, and associations between measured potassium and arrhythmia do not establish the effect of giving additional potassium.¹

TIGHT K compared supplementation triggers of 3.6 and 4.5 mEq/L after isolated coronary artery bypass grafting (CABG). It asked whether the less intensive policy retained acceptable effectiveness while reducing unnecessary treatment. The study used a non-inferiority design because a strategy requiring fewer interventions might be worthwhile without preventing more arrhythmias. Its results therefore need to be interpreted against the chosen margin, the achieved potassium separation and the quality of rhythm ascertainment. They also concern prevention in patients without a history of atrial dysrhythmia, rather than treatment of an established arrhythmia or permission to leave severe hypokalaemia uncorrected.¹

Synopsis

TIGHT K was an open-label randomised trial in 23 cardiac surgical centres, 21 in the United Kingdom and two in Germany. Recruitment occurred from October 2020 to November 2023. Adults in sinus rhythm who were scheduled for isolated CABG were eligible. Additional cardiac or vascular procedures excluded participation. Other exclusions included previous atrial fibrillation, flutter or atrial tachyarrhythmia, high-degree atrioventricular block, medication used for rhythm management, preoperative potassium above 5.5 mEq/L and dialysis-dependent kidney failure. Consent was obtained before surgery.^{1,2}

Of 5,568 patients assessed, 1,690 were randomised. After removal of three randomisations made in error, the safety population comprised 843 patients assigned to relaxed supplementation and 844 assigned to tight supplementation. Further exclusions before intervention or after identification of ineligibility left an efficacy population of 830 and 837. These included patients who did not undergo isolated CABG, one intraoperative death and one withdrawal. Allocation used a secure central service, variable blocks and stratification by site. Patients and caregivers knew the assignment, while the core rhythm-monitor laboratory and event-validation committee were masked.¹

The assigned policy began on arrival in the postoperative care facility, provided the patient was in sinus or paced rhythm. Potassium was supplemented only when its measured concentration fell below 3.6 mEq/L in the relaxed group or below 4.5 mEq/L in the tight group. Sampling frequency, point-of-care or laboratory testing, route of replacement and other care followed local practice. Magnesium, β -blockers and amiodarone were not prohibited. Assigned treatment ended at 120 hours, hospital discharge or the first site-reported atrial arrhythmia, whichever occurred first. Subsequent potassium management was unrestricted.¹

An ambulatory heart rhythm monitor supplemented routine monitoring for the first 120 postoperative hours or until discharge. The primary endpoint was described as clinically detected and electrocardiographically confirmed new-onset atrial fibrillation after cardiac surgery (AFACS). Its definition also included atrial flutter and atrial tachyarrhythmia lasting at least 30 seconds or present throughout an entire 12-lead recording. The independent committee reviewed reported events using ECG or telemetry, ambulatory recordings and, if necessary, supporting clinical documentation. The supplemental adjudication rules also permitted probable events to contribute when sufficient confirmatory evidence was unavailable.^{1,2}

Secondary outcomes included AFACS detected by the ambulatory monitor, AFACS identified by either method, other dysrhythmias, inpatient mortality, critical care and hospital stay, and potassium purchase and administration costs. Arrhythmia burden was exploratory. The trial's non-inferiority margin was an absolute increase in clinically detected AFACS of ten percentage points. Non-inferiority required the upper limit of a one-sided 97.5% confidence interval to remain below that margin. The planned recruitment target was 1,684, allowing for losses, based on 90% power, an expected 35% event rate with tight control and an assumed two-percentage-point excess with relaxed control.¹

The primary analysis used the efficacy population without covariate adjustment. An additional analysis adjusted for age, sex and site. A per-protocol analysis excluded patients with specified treatment deviations: repeated inappropriate supplementation in the relaxed group or prolonged failure to supplement below the trigger in the tight group. This excluded 48 relaxed-group patients and 135 tight-group patients, leaving 782 and 702. Analyses of monitor-based outcomes also used inverse probability weighting to examine the effect of missing recordings. Secondary clinical comparisons were superiority analyses.^{1,2}

Mean age was approximately 65 years, 15% were women and most participants were White. Mean EuroSCORE II was about 1.5%, indicating a generally low predicted operative mortality. Approximately 86% underwent cardiopulmonary bypass, 78% received a β -blocker at baseline, and mean baseline potassium was 4.4 mEq/L. Subsequent magnesium and antiarrhythmic treatment were common. The treatment policies produced clear separation: the median number of potassium administrations during the assigned period was zero in the relaxed group and seven in the tight group. Mean potassium concentrations remained higher with tight supplementation, while measurement frequency was similar.^{1,2}

The primary endpoint occurred in 231 of 830 relaxed-group patients and 219 of 837 tight-group patients, or 27.8% versus 26.2%. The unadjusted difference was 1.7 percentage points (95% CI -2.6 to 5.9). Its upper limit satisfied the ten-percentage-point non-inferiority margin. The adjusted difference was 2.2 percentage points (95% CI -1.9 to 6.4). In the per-protocol population, events occurred in 221 of 782 versus 202 of 702, or 28.3% versus 28.8%, with a difference of -0.5 percentage points (95% CI -5.1 to 4.1). These analyses supported the primary non-inferiority finding.^{1,2}

AFACS identified clinically or by ambulatory monitoring occurred in 275 patients with relaxed control and 276 with tight control, or approximately 33% in each group. Among patients with adequate monitor data, monitor-detected AFACS occurred in 32.2% versus

33.1%. Other monitor-detected dysrhythmias occurred in 19.1% versus 21.1%, without an established difference. There were four inpatient deaths in each efficacy group. Median critical care stay was two days in both groups, and median hospital stay was six days in both.^{1,2}

The reported mean cost of potassium purchase and administration was £30.63 per patient with relaxed control and £117.83 with tight control. The estimated saving was £87.21 (95% CI £80.74 to £93.67), equivalent to the report's US-dollar estimate of \$111.89 using its historical conversion method. This was a focused costing of potassium treatment, incorporating drugs, staff time and consumables, rather than a full economic evaluation. The trial supported a lower supplementation trigger within its specified population and non-inferiority margin, with substantially fewer administrations and lower treatment costs.^{1,2}

Critique

TIGHT K addresses a specific, common and readily changeable practice. Its large multicentre sample, central randomisation and clear contrast in supplementation make the result directly useful. A median of zero versus seven administrations demonstrates that the intervention altered care substantially. The accompanying potassium separation also supports a genuine physiological contrast. The result cannot readily be dismissed as a comparison of two effectively identical treatment policies, although the difference in serum concentrations was considerably smaller than the difference between their trigger thresholds.¹

That distinction is essential for application. A trigger is the concentration below which treatment is given; it is not a target around which every patient must be maintained. In Figure 3, mean potassium in the relaxed group remained above 4.0 mEq/L throughout the reported periods, despite the 3.6 mEq/L trigger. Many patients did not require supplementation. The study therefore does not show that deliberately maintaining potassium near 3.6 mEq/L is as safe as maintaining it near 4.5 mEq/L. It supports avoiding routine supplementation of an otherwise monitored population whose concentrations commonly remain within the normal range without it.¹

The non-inferiority margin was generous. An allowable ten-percentage-point excess would represent a substantial relative increase from the observed tight-group risk of 26.2%. The actual confidence interval was narrower and its upper limit of 5.9 percentage points comfortably met the trial's criterion, but still permits a smaller clinically meaningful increase. Thus, the result establishes non-inferiority under the selected margin rather than identical risks. The non-significant superiority P values shown

alongside the outcomes are not the basis for that conclusion. Whether the remaining uncertainty is acceptable must be weighed against the clear reduction in treatment burden.¹

Concordance between efficacy and per-protocol analyses is reassuring, because nonadherence can make treatment groups appear more similar in a non-inferiority trial. Nevertheless, departures were asymmetric: approximately 16% of tight-group patients and 6% of relaxed-group patients were excluded from the per-protocol analysis. Tight-group staff sometimes did not supplement when potassium was only slightly below 4.5 mEq/L. The selected adherent groups also had different event rates from the excluded patients. Per-protocol analysis therefore provides supportive evidence, without recreating an unbiased randomised comparison of perfect adherence. The strong administration contrast in the main analysis remains especially informative.^{1,2}

The distinction between clinical and monitor-detected arrhythmia is another important strength and limitation. Clinical detection captures events recognised during care, but staff knowledge of allocation might influence attention, investigation or reporting. Masked laboratory analysis and independent adjudication help address this problem. However, the adjudication hierarchy allowed supporting documentation and probable events when recordings were unavailable or inconclusive. The primary outcome should therefore not be interpreted as requiring an independently reviewable electrocardiogram in every case. The report does not provide a clear breakdown showing how many counted events depended on the least certain evidence categories.^{1,2}

Ambulatory recording was also incomplete. The monitor-based AFACS analysis lacked adequate data for 147 relaxed-group patients and 133 tight-group patients, approximately 17% of the efficacy population overall. Other dysrhythmia analyses had further missing data. The outcome combining clinical and monitor detection retained the full efficacy denominator, but could miss otherwise silent events in patients without usable monitoring. Inverse probability weighting produced similar estimates, which is useful reassurance under its assumptions; it cannot recover unobserved rhythm information or establish that missing recordings were unrelated to events after accounting for measured variables.^{1,2}

The absence of an established difference in secondary arrhythmias or mortality does not constitute a separate non-inferiority demonstration for every safety outcome. Only eight inpatient deaths occurred in the efficacy population, so substantial relative mortality differences remain compatible with the data. The ventricular tachycardia category included runs of three or more beats, and should not be read as a count of sustained life-threatening arrhythmias. The one ventricular fibrillation event described in the

supplement occurred before assigned treatment began, with potassium of 4.7 mEq/L. It provides no evidence that the relaxed trigger caused hypokalaemic ventricular fibrillation.^{1,2}

Treatment stopped after the first site-reported AFACS episode. That is sensible for a prevention trial, because clinicians then needed freedom to treat the arrhythmia. It also means that subsequent arrhythmia burden, hospital outcomes and costs do not represent uninterrupted exposure to the assigned potassium policy in every patient. The study primarily answers how to supplement before the first postoperative episode. It does not establish the best potassium threshold during active AFACS, recurrent arrhythmia or treatment with drugs whose safety depends on electrolyte status.¹

Background treatment and the recruited population further define the result. β -blockers were common, magnesium was administered to roughly three quarters of participants, and amiodarone was used in approximately one fifth during follow-up. These treatments were not randomised, and some may have followed an arrhythmia rather than prevented one. Subgroup analyses did not establish treatment-effect heterogeneity, but several intervals were wide. In particular, the apparent signal among patients without baseline β -blockers cannot validate a separate supplementation policy. Women, patients with poor ventricular function and those at higher predicted operative risk were represented in smaller numbers. Dialysis patients and those having combined valve procedures were excluded.^{1,2}

Finally, the financial finding is a treatment-cost estimate derived largely from UK unit costs and assumed nursing time. It supports reduced resource consumption, but is not a demonstrated reduction in total hospital expenditure or a complete cost-effectiveness analysis. The listed route-specific cost means also do not transparently reconcile with the overall mean, so the reported total and its uncertainty should be distinguished from an independently reconstructed costing. Likewise, fewer administrations plausibly reduce inconvenience and opportunities for medication error, but the trial did not prove reductions in every proposed downstream harm. Its strongest conclusion is that routine high-normal supplementation can be reduced in monitored isolated-CABG patients within the limits studied.^{1,2}

Body of Evidence

The direct precursor was the TIGHT K pilot and feasibility study, which recruited 160 patients in two hospitals. It tested whether patients undergoing first-time elective isolated CABG could be assigned to the same two potassium policies. Recruitment, treatment adherence and follow-up were the main feasibility objectives. The reported violation rate was 9.8% and 28-day follow-up was 94.3%. These findings supported conducting the definitive trial; they should not be presented as an independently powered demonstration that the two strategies had equivalent clinical outcomes.³

Observational evidence had offered conflicting signals. Howitt and colleagues analysed 3,068 cardiac surgical patients, of whom 545 developed AF within 72 hours. A mean potassium concentration below 4.5 mmol/L was associated with higher odds of AF after adjustment (odds ratio 1.43; 95% CI 1.17 to 1.75). This association helped sustain the rationale for high-normal supplementation, but could not determine the effect of administering potassium. Illness severity, treatment responses and the timing of measurements can influence both the recorded concentration and arrhythmia risk.⁴

Lancaster and colleagues reached a different observational conclusion in 2,041 patients undergoing CABG, valve surgery or both. Patients developing postoperative AF had higher potassium concentrations at the time of onset than time-matched controls, and prophylactic potassium supplementation was not associated with a lower AF rate. The study reduced some timing problems by matching measurements, but treatment was still not randomly assigned. These conflicting associations illustrate why neither a low measured concentration nor a high one should automatically be interpreted as the causal explanation for postoperative arrhythmia.⁵

The broader importance of postoperative AF is supported by a systematic review from Caldonazo and colleagues, incorporating 57 studies and 246,340 patients. AF after cardiac surgery was associated with adverse perioperative and longer-term outcomes, including mortality and stroke. These associations make prevention clinically relevant, but do not establish that every intervention reducing recorded AF will reduce those outcomes. Patients developing AF can differ in underlying vulnerability and perioperative illness. TIGHT K therefore needed to test the supplementation policy directly, rather than rely on the prognostic importance of the arrhythmia as proof of benefit.⁶

PAPABEAR demonstrates that prevention can be evaluated using a distinct pharmacological intervention. In 601 patients scheduled for non-emergency CABG or valve surgery, a brief perioperative course of oral amiodarone reduced atrial tachyarrhythmias lasting at least five minutes and prompting treatment by

postoperative day six: 16.1% versus 29.5%. Dose reductions were more frequent with amiodarone, and mortality was not demonstrably reduced. Its eligibility, drug exposure and endpoint differed from TIGHT K, so it does not answer the potassium question. It does show why evidence for an active antiarrhythmic strategy cannot be assumed to apply to electrolyte supplementation.⁷

Against this background, TIGHT K supplies the strongest direct randomised evidence for the specific decision it examined. The finding is consistent with the possibility that correcting genuine potassium deficiency matters while additional supplementation within the normal range contributes little to AFACS prevention in many isolated-CABG patients. It does not identify a universal optimal potassium concentration across critical care or cardiac surgery. Nor does it test withdrawal of rhythm surveillance, management of ongoing electrolyte losses, or cessation of other preventive treatments. The useful change is a more selective trigger for supplementation, with the clinical monitoring and exclusions that defined the trial kept in view.¹

Can the potassium supplementation trigger be relaxed after isolated CABG in patients without previous atrial dysrhythmia?

Yes: supplementation below 3.6 mEq/L met TIGHT K's non-inferiority criterion compared with a 4.5 mEq/L trigger and reduced treatment burden. This supports continued monitoring with selective replacement, within the population studied.

References

1. O'Brien B, Campbell NG, Allen E, Jamal Z, Sturgess J, Sanders J, et al.; TIGHT K investigators. Potassium supplementation and prevention of atrial fibrillation after cardiac surgery: the TIGHT K randomized clinical trial. *JAMA* 2024;332(12):979–988.
2. TIGHT K investigators. Supplement 3 to: Potassium supplementation and prevention of atrial fibrillation after cardiac surgery: the TIGHT K randomized clinical trial. *JAMA* 2024;332(12):979–988.
3. Campbell NG, Allen E, Montgomery H, Aron J, Canter RR, Dodd M, et al. Maintenance of serum potassium levels ≥ 3.6 mEq/L versus ≥ 4.5 mEq/L after isolated elective coronary artery bypass grafting and the incidence of new-onset atrial fibrillation: pilot and feasibility study results. *J Cardiothorac Vasc Anesth* 2022;36(3):847–854.

4. Howitt SH, Grant SW, Campbell NG, Malagon I, McCollum C. Are serum potassium and magnesium levels associated with atrial fibrillation after cardiac surgery? *J Cardiothorac Vasc Anesth* 2020;34(5):1152–1159.
5. Lancaster TS, Schill MR, Greenberg JW, Moon MR, Schuessler RB, Damiano RJ, et al. Potassium and magnesium supplementation do not protect against atrial fibrillation after cardiac operation: a time-matched analysis. *Ann Thorac Surg* 2016;102(4):1181–1188.
6. Caldonazo T, Kirov H, Rahouma M, Robinson NB, Demetres M, Gaudino M, et al.; POAF-MA Group. Atrial fibrillation after cardiac surgery: a systematic review and meta-analysis. *J Thorac Cardiovasc Surg* 2023;165(1):94–103.e24.
7. Mitchell LB, Exner DV, Wyse DG, Connolly CJ, Prystai GD, Bayes AJ, et al. Prophylactic oral amiodarone for the prevention of arrhythmias that begin early after revascularization, valve replacement, or repair: PAPABEAR: a randomized controlled trial. *JAMA* 2005;294(24):3093–3100.

Possible Reporting Inconsistencies. The discussion states that mean potassium was not above each group's trigger, whereas Figure 3 shows relaxed-group means above 4.0 mEq/L, exceeding its 3.6 mEq/L trigger. In eTable 3, 820 tight-group β -blocker recipients plus 35 non-recipients exceed the stated group size of 837. Route-specific cost means in Table 2 and eTable 10 do not add to the overall means; their denominators need clarification. The chapter uses the reported total costs and primary-outcome estimates.



Miscellaneous Trials

DEFENDER

Tavares CAM, Azevedo LCP, Rea-Neto Á, Campos NS, Amendola CP, Kozesinski-Nakatani AC, et al.; DEFENDER Investigators. Dapagliflozin for critically ill patients with acute organ dysfunction: the DEFENDER randomized clinical trial. JAMA 2024;332(5):401–411.

Introduction

Treatments that protect the heart and kidney during chronic disease are attractive candidates for critical illness. Sodium–glucose cotransporter 2 inhibitors offer a particularly interesting possibility: their clinical effects extend beyond glucose lowering, suggesting that patients without diabetes might also benefit. However, preventing deterioration over years and reversing organ dysfunction over days are substantially different therapeutic tasks. Evidence from stable outpatients cannot establish that starting the same medication during shock, respiratory failure or acute kidney injury improves recovery.¹

The proposed rationale includes changes in renal haemodynamics, tubular workload, cellular energy handling and inflammatory signalling. These mechanisms might reduce the metabolic cost of maintaining organ function during acute illness. They remain hypotheses in this setting, and neither a plausible pathway nor an established benefit in chronic disease guarantees a useful response in critical care. Enteral absorption, the short treatment window and the underlying cause of organ dysfunction could all alter effectiveness.^{1,2}

Safety also requires direct investigation. A drug that increases urinary glucose losses is being introduced into patients whose nutrition, insulin requirements, circulation and renal function may change rapidly. Ketoacidosis can occur without marked hyperglycaemia, making routine glucose measurement an incomplete safeguard. Equally, assuming that these concerns make carefully supervised investigation impossible would prevent the question from being answered.¹

DEFENDER tested dapagliflozin early after the development of acute organ dysfunction in a broad Brazilian critical care population. It combined a clinically ordered primary outcome with separate analyses of survival, kidney replacement therapy and time free of organ support. The resulting distinction is central to interpretation: the overall trial was neutral, while one kidney outcome generated an exploratory signal that warrants scrutiny rather than immediate adoption.¹

Synopsis

DEFENDER was an investigator-initiated, open-label randomised trial in 22 Brazilian intensive care units. Between November 2022 and August 2023, it enrolled 507 adults with an unplanned admission, an anticipated critical care stay of at least 48 hours and at least one recently developed respiratory, cardiovascular or kidney dysfunction. The relevant dysfunction had to have been present for no more than 24 hours.^{1,3}

Eligibility included hypotension or vasopressor treatment, a creatinine increase of at least 0.3 mg/dL or sustained oliguria, or a new requirement for high-flow nasal oxygen, non-invasive ventilation or invasive ventilation. These were pragmatic organ dysfunction criteria, rather than a requirement for sepsis or established multiorgan failure. Exclusions included maintenance dialysis, planned admission after elective surgery, previous sodium–glucose cotransporter 2 inhibitor use, type 1 diabetes, previous diabetic ketoacidosis, pregnancy and total fasting or inability to receive enteral medication.^{1,3}

Concealed web-based randomisation allocated 248 patients to dapagliflozin and 259 to standard care, using variable blocks within sites. Dapagliflozin was given at 10 mg once daily, orally or crushed through an enteral tube, starting within 24 hours of allocation. Treatment continued for up to 14 days or until critical care discharge, whichever occurred first. Local teams otherwise directed care. Both groups were advised to receive at least 100 g of carbohydrate daily.¹

Dapagliflozin was withheld for specified clinical circumstances, including inability to receive enteral treatment, ketoacidosis, repeated severe hypoglycaemia and initiation of kidney replacement therapy. All allocated patients received at least one dose, and no control patient received the drug. Thirteen intervention patients missed at least one scheduled dose. Exposure nevertheless shortened as patients left critical care or died: the administration figure shows 129 of 248 receiving dapagliflozin on day 5 and 27 on day 14. These proportions describe the original allocation group, not adherence among patients still eligible for treatment.^{1,3}

Participants had a mean age of approximately 64 years, 47% were women and about one-third had type 2 diabetes. Around 40% had suspected infection, 46% were invasively ventilated and half were receiving noradrenaline at baseline. Most admissions were medical, with the remainder following emergency surgery. Randomisation occurred a median of one day after critical care admission. Patients could satisfy more than one organ dysfunction criterion, although many had only one qualifying dysfunction.¹

The primary outcome used a hierarchy of hospital mortality, new kidney replacement therapy and cumulative critical care length of stay through 28 days. Each intervention patient was compared with every control patient. Survival decided a comparison when only one patient died; if both died, the pair tied without considering lower outcomes. Among survivors, avoiding kidney replacement therapy took priority, followed, where necessary, by a shorter critical care stay. The win ratio divided intervention wins by control wins. It was not the proportion of patients experiencing a conventional binary composite.^{1,3}

The planned sample of approximately 500 provided at least 85% power under simulations combining several benefits: mortality falling from 30% to 28%, kidney replacement therapy from 10% to 7%, and a modest reduction in length of stay. These assumptions produced an anticipated win ratio around 1.40. Secondary binary and ordinal outcomes were analysed using Bayesian models adjusted for site and baseline clinical characteristics, with a neutral treatment prior. Its central 95% interval covered odds ratios from 0.5 to 2.0. Clinical adjustment included age, suspected sepsis, vasopressor use and invasive ventilation at randomisation. Complementary frequentist analyses and exploratory competing-risk analyses were also reported.^{1,3}

The primary analysis included all 507 patients. Across 64,232 comparisons, dapagliflozin produced 27,143 wins and control 26,929, giving a win ratio of 1.01 (95% CI 0.90–1.13; $P=0.89$). Vital status was known for everyone. Three transferred control patients lacked complete subsequent organ-support and length-of-stay information; their lower-level primary comparisons were truncated at transfer. They were excluded from the corresponding secondary analyses, explaining why the kidney replacement therapy denominator was 256 rather than 259.^{1,3}

Hospital mortality was 88/248 (35.5%) with dapagliflozin and 89/259 (34.4%) with control. The adjusted odds ratio was 1.06 (95% credible interval 0.76–1.52), with a posterior probability of benefit of 0.36. Kidney replacement therapy occurred in 27/248 (10.9%) and 39/256 (15.2%): adjusted odds ratio 0.76 (95% credible interval 0.51–1.18), adjusted absolute difference –2.9 percentage points (–7.4 to 1.7), and posterior probability of benefit 0.90.^{1,3}

Analyses of days alive and free of critical care, hospitalisation, ventilation, vasopressors or kidney replacement therapy were inconclusive. Death before discharge attracted the worst score; the calendar-day scale ran from zero to 29. Ventilation and vasopressor use for six hours or less in a calendar day qualified that day as free of the corresponding support. A post hoc modified major adverse kidney event composite combined death, kidney replacement therapy and creatinine doubling, with creatinine available only

during the first five study days. Its adjusted odds ratio was 0.96 (95% credible interval 0.70–1.33). Competing-risk estimates for kidney replacement therapy also remained uncertain.^{1,3}

Serious adverse events were reported in 115/248 (46.4%) and 123/259 (47.5%). There were no reported ketoacidosis events. Severe hypoglycaemia occurred in two intervention patients and none receiving control; urinary tract infection occurred in four and three, respectively. Serial creatinine and pH measurements over the first five days did not show convincing between-group differences. The investigators concluded that dapagliflozin had not improved overall clinical outcomes, while acknowledging uncertainty about benefits and harms for individual outcomes.^{1,3}

Critique

DEFENDER addressed a new therapeutic question in patients with substantial acute illness, rather than relying on extrapolation from outpatient trials. Its multicentre recruitment, concealed allocation, complete ascertainment of hospital vital status and retention of all randomised patients in the primary analysis are strengths. The substantial use of ventilation and vasopressors makes the study directly relevant to critical care. Public funding through the Brazilian Ministry of Health was reported, with the funder having no role in trial conduct, analysis or the publication decision.¹

The hierarchy gives death precedence over less important outcomes within each pair. That is a sensible response to a composite in which discharge timing might otherwise compete indiscriminately with survival. Nevertheless, the summary measure remains dependent on the selected hierarchy and comparison rules. Death decided 29,200 comparisons; critical care length of stay decided another 20,010. Giving mortality first priority does not mean that mortality alone determines the treatment estimate. The 64,232 comparisons also do not constitute 64,232 independent patients or an equivalently large trial.^{1,3}

A win ratio of 1.01 indicates almost identical numbers of wins under these rules. It cannot be translated into a 1% improvement in survival, kidney recovery or the chance that a particular patient benefits. Its confidence interval excludes the large overall effect used in the sample-size simulations, but the intervals for individual hard outcomes remain considerably wider. The trial therefore argues against the anticipated combined benefit while leaving clinically relevant uncertainty about mortality and kidney replacement therapy. It was neither a non-inferiority trial nor proof of therapeutic equivalence.¹

Open treatment allocation matters most for the lower components. Kidney replacement therapy initiation and critical care discharge depend partly on clinician judgement and local practice. Awareness of treatment could influence those decisions even without intentional preferential care. Site adjustment in secondary models accommodates some baseline differences between units, but cannot remove differential behaviour after allocation. Mortality is less susceptible to this mechanism, and it showed no favourable numerical pattern. The neutral aggregate result is consequently not rescued by selecting the most encouraging discretionary component.^{1,3}

The Bayesian kidney finding deserves accurate language. A 0.90 probability of benefit is conditional on the data, model and prior; it is not a 90% probability that the whole trial succeeded. Nor is it a frequentist P value of 0.10. The credible interval spans reduced and increased odds of treatment, and the analysis belongs among several secondary outcomes. Its value is to identify residual uncertainty and a potential research direction. It does not supply confirmatory evidence that dapagliflozin prevents dialysis.^{1,3}

Competing death is particularly relevant: patients who die early have less opportunity to start kidney replacement therapy. The exploratory cause-specific and subdistribution hazard models both favoured dapagliflozin numerically, but their confidence intervals included no effect. The broader modified kidney composite was also unconvincing. These analyses do not independently replicate the secondary finding; they interrogate the same small dataset with related outcomes. Furthermore, creatinine doubling during five days is not equivalent to persistent renal dysfunction at 30 or 90 days.^{1,3}

The handling of free days also affects clinical meaning. Patients discharged before day 28 were assumed to remain alive and free of support thereafter; follow-up was not a comprehensive assessment of all deaths and readmissions after discharge. Reported median hospital-free days of 8.5 and 2.5 should therefore not be described as a proven six-day treatment benefit. A difference between group medians is not the adjusted treatment estimate, and distributions with many deaths at zero can yield striking median changes without a convincing overall shift.^{1,3}

The three incomplete transfers are a small fraction of recruitment, but their management should remain visible. Knowing that a patient survived does not establish their subsequent dialysis use or critical care stay. Truncating comparisons at transfer preserved inclusion in the primary analysis, while excluding incomplete secondary observations changed those denominators. The supplement explains the distinction. Presenting all secondary analyses as complete intention-to-treat follow-up would conceal it, even though three patients are unlikely to transform the broad neutral interpretation.³

Biological exposure is a further uncertainty. Receipt of an enteral dose does not demonstrate absorption during acute illness, and the investigators did not measure drug concentrations or a specific biological response. Most participants also received substantially less than 14 days of treatment. These facts limit mechanistic conclusions; they do not invalidate the clinical test of this early, discharge-limited policy. Extending treatment beyond critical care or selecting patients by measured target engagement would constitute a different intervention requiring its own evaluation.^{1,3}

The reassuring safety observations need the same boundaries. This was a selected population able to receive enteral medication, with nutritional advice, laboratory monitoring and rules for stopping treatment. It excluded important groups at greater risk of ketoacidosis and discontinued dapagliflozin when kidney replacement therapy began. No reported ketoacidosis among 248 treated patients cannot establish that a rare event is absent, and comparable overall adverse-event proportions are not a formal demonstration of equal safety. The results support further monitored investigation, rather than unrestricted initiation during critical illness.^{1,3}

Finally, heterogeneity offers a possible explanation for a modest average effect, but not an evidence-based way to identify responders. Analyses by diabetes, suspected sepsis, creatinine, age and cardiovascular admission did not establish convincing treatment interactions. In particular, a baseline creatinine category does not reliably distinguish acute injury from chronic impairment. A more focused kidney trial could be reasonable, provided its population, renal outcomes and follow-up are chosen prospectively. DEFENDER itself supplies no validated subgroup for routine treatment.^{1,2}

Body of Evidence

The strongest rationale for renal protection came from chronic disease. DAPA-CKD randomised 4,304 participants with reduced estimated glomerular filtration and substantial albuminuria to dapagliflozin or placebo. Over a median 2.4 years, the composite of sustained major loss of kidney function, end-stage kidney disease, or renal or cardiovascular death occurred in 9.2% versus 14.5% (hazard ratio 0.61, 95% CI 0.51–0.72). Benefit was observed with and without diabetes. This established an important disease-modifying role in that population, but the duration, selection and outcome definitions differ fundamentally from a short course during established acute organ dysfunction. DEFENDER tested a new indication, rather than repeating DAPA-CKD under more severe conditions.⁴

DARE-19 moved the question into acute hospital care. It randomised 1,250 patients hospitalised with COVID-19 and at least one cardiometabolic risk factor to dapagliflozin

or placebo for 30 days, excluding those critically ill at screening. Organ dysfunction or death occurred in 11.2% versus 13.8% (hazard ratio 0.80, 95% CI 0.58–1.10; $P=0.17$). Its other primary outcome, a hierarchical measure of clinical recovery, was also inconclusive: win ratio 1.09 (95% CI 0.97–1.22). Numerical trends supported continued investigation, but neither primary outcome demonstrated benefit. Its lower-risk, COVID-specific population cannot establish efficacy in patients already requiring critical care organ support.⁵

RECOVERY provided a larger acute-illness test with another drug in the class. Among 4,271 adults hospitalised with COVID-19, empagliflozin plus usual care produced 289/2,113 deaths by day 28, compared with 307/2,158 receiving usual care: rate ratio 0.96 (95% CI 0.82–1.13). Hospital stay and progression to invasive ventilation or death were also not convincingly improved. Two treatment-related serious adverse events involved ketosis without acidosis. This trial makes a substantial, general survival benefit in hospitalised COVID-19 less plausible. It does not directly settle dapagliflozin treatment in non-COVID critical illness, but it limits the argument that promising biological effects already translate into broad acute organ protection.⁶

EMPULSE illustrates a setting in which starting treatment during hospitalisation did show benefit. It enrolled 530 patients with acute heart failure after clinical stabilisation, a median of three days after admission. Empagliflozin improved a 90-day hierarchical outcome incorporating death, heart failure events and symptoms: stratified win ratio 1.36 (95% CI 1.09–1.68). Patients were treated for up to 90 days, and the hierarchy included a patient-reported symptom score. These differences matter. A favourable result after stabilisation for a defined cardiac syndrome cannot be assumed to apply during heterogeneous acute organ dysfunction, and win ratios from differently constructed hierarchies are not directly comparable measures of biological potency.⁷

Together, these trials show why the clinical setting matters more than the broad label of organ protection. Chronic kidney disease, stabilised acute heart failure, hospitalised COVID-19 and newly established critical care organ dysfunction pose different treatment questions. Positive chronic or cardiac evidence sustains biological interest; the neutral acute infection trials and DEFENDER constrain clinical extrapolation. None provides independent confirmation of DEFENDER's secondary kidney replacement therapy finding.^{1,4,5,6,7}

As of the end of 2024, DEFENDER therefore supported equipoise for a carefully designed further trial, especially one able to characterise renal injury and recovery over an appropriate interval. Measuring exposure and a relevant biological response alongside clinical outcomes could help distinguish failure to deliver an active treatment

from failure of the proposed therapeutic mechanism. It did not establish benefit from routinely starting dapagliflozin for acute organ dysfunction. It also did not answer whether to continue or restart a pre-existing, independently indicated prescription, because previous use was an exclusion. That separate decision should not be represented as a result of this trial.^{1,3}

Should dapagliflozin be started routinely for acute organ dysfunction in critical care?

No. DEFENDER found no overall clinical benefit; its uncertain kidney replacement therapy signal supports further research, while its safety observations apply to selected, monitored patients receiving enteral treatment.

References

1. Tavares CAM, Azevedo LCP, Rea-Neto Á, Campos NS, Amendola CP, Kozesinski-Nakatani AC, et al.; DEFENDER Investigators. Dapagliflozin for critically ill patients with acute organ dysfunction: the DEFENDER randomized clinical trial. JAMA 2024;332(5):401–411.
2. Gómez H, Derde LPG. Sodium-glucose cotransporter 2 therapy for acute organ dysfunction in critically ill patients. JAMA 2024;332(5):377–379.
3. DEFENDER Investigators. Supplement 3 to: Dapagliflozin for critically ill patients with acute organ dysfunction: the DEFENDER randomized clinical trial. JAMA 2024;332(5):401–411.
4. Heerspink HJL, Stefánsson BV, Correa-Rotter R, Chertow GM, Greene T, Hou FF, et al.; DAPA-CKD Trial Committees and Investigators. Dapagliflozin in patients with chronic kidney disease. N Engl J Med 2020;383(15):1436–1446.
5. Kosiborod MN, Esterline R, Furtado RHM, Oscarsson J, Gasparyan SB, Koch GG, et al. Dapagliflozin in patients with cardiometabolic risk factors hospitalised with COVID-19 (DARE-19): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Diabetes Endocrinol 2021;9(9):586–594.
6. RECOVERY Collaborative Group. Empagliflozin in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. Lancet Diabetes Endocrinol 2023;11(12):905–914.

7. Voors AA, Angermann CE, Teerlink JR, Collins SP, Kosiborod M, Biegus J, et al. The SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure: a multinational randomized trial. *Nat Med* 2022;28(3):568–574.

Possible Reporting Inconsistencies. The accompanying editorial labels 0.90 as a P value for kidney replacement therapy, whereas the trial reports a Bayesian posterior probability of benefit; its description of that analysis as post hoc also differs from the index report. In the main safety table, 15/248 nervous system events are printed as 0.6%; the detailed supplement reports 6.0%. Supplementary Table 5 gives control vasopressor-free days as a median of 25 with an IQR of 17–22, which cannot contain that median. These presentation discrepancies do not alter the neutral primary result.

TELESCOPE

Pereira AJ, Noritomi DT, dos Santos MC, Corrêa TD, Ferraz LJR, Schettino GPP, et al. Effect of tele-ICU on clinical outcomes of critically ill patients: the TELESCOPE randomized clinical trial. JAMA 2024;332(21):1798–1807.

Introduction

Access to specialist advice is unevenly distributed across critical care. A hospital may have beds, ventilators and an on-site clinical team while lacking regular input from an intensivist. Telemedicine offers a way to connect those teams with expertise elsewhere, potentially supporting diagnosis, treatment planning and reliable delivery of established care. The attraction is clear, but making expertise available is only the first step towards changing what happens to patients.¹

A remote recommendation must be relevant, communicated clearly, accepted by the local team and feasible with the resources available. It must then be implemented at the right time and produce a meaningful change in care. These stages depend on relationships, staffing, workload and local authority as much as on video quality or electronic connectivity. A technically successful consultation can therefore coexist with no measurable improvement in clinical outcomes.^{1,2}

Evaluating telemedicine is also difficult because the term encompasses substantially different services. Scheduled specialist rounds, continuous surveillance, emergency consultation and remote authority to initiate treatment are not interchangeable. Before-and-after evaluations can compound this problem: the introduction of a service may coincide with wider staffing or quality improvements, creating an apparent benefit that cannot confidently be attributed to telemedicine itself.^{1,2}

TELESCOPE addressed this uncertainty with a parallel cluster randomised trial in Brazilian public hospitals. It tested a programme combining weekday multidisciplinary tele-rounds, performance feedback and clinical guidance in units without established specialist-led rounds. Its principal question was whether this package shortened critical care stay. The neutral result is informative precisely because the study evaluated an identifiable service in a defined organisational setting, rather than telemedicine as an abstract idea.¹

Synopsis

Thirty general adult intensive care units, one per hospital, participated across Brazil's geographic regions. Units required at least eight beds and on-site doctors and nurses. Specialist units, private hospitals and units with established structured multidisciplinary rounds or regular management through performance indicators were excluded. The selected hospitals lacked the full-time specialist bedside coverage that the programme aimed to supplement. Individual eligibility was broad, including consecutive adults except those whose admission or discharge was determined by exceptional legal rather than medical circumstances. At baseline, 20 units reported no multidisciplinary rounds, and the other ten described irregular or unstructured rounds. Twenty-four units reported no audit and feedback using performance indicators, identifying substantial room to improve these organisational processes.¹

Each unit contributed a two-month baseline period. A concealed central allocation then assigned 15 units to telemedicine and 15 to usual care, using a restricted algorithm to balance characteristics including bed numbers, illness severity, length of stay, resource use, mortality and region. Randomisation occurred in three batches for practical reasons. Units retained their assigned group; this was not a stepped-wedge design in which all controls subsequently received the intervention. Patients admitted while units awaited allocation were excluded.^{1,3}

Recruitment ran from June 2019 to April 2021, with in-hospital follow-up completed in July. There were 1,794 patients in the baseline period and 15,230 in the intervention period: 7,471 in telemedicine units and 7,759 in usual-care units. All 30 units contributed to the analysis. Consent for individual patients was waived, and trained personnel outside the clinical care team collected data. Follow-up was limited to the hospital episode, with a maximum of 90 days after critical care admission.¹

The programme had three components. A remote specialist intensivist conducted structured bedside video rounds with local staff from Monday to Friday. Monthly meetings with local leaders reviewed performance indicators and improvement plans. Nineteen sets of clinical guidance were supplied through text and video materials. The same remote intensivist maintained continuity with each unit, had training in communication and quality improvement, and was available for consultation outside the scheduled rounds. Equipment and software were provided without charge to participating units.^{1,3}

Doctors and nurses were required at the rounds, with physiotherapy participation strongly encouraged. The remote clinician reviewed active problems and proposed a plan until the next round. Local doctors retained responsibility for diagnostic and

treatment decisions: the service did not directly prescribe medication or order investigations. Teams were asked to record acceptance of recommendations within 24 hours. Usual-care units received no tele-rounds or audit meetings, although both groups received COVID-19 guidance and training materials after the pandemic began.^{1,3}

The primary outcome was the first critical care admission's length of stay, measured from admission until discharge, transfer or death. A linear mixed model analysed log-transformed duration, accounting for baseline data, clustering and patient and unit characteristics. Its treatment estimate was expressed as a percentage change. The initial target was 15,000 intervention-period patients across 30 units, intended to detect a 1.5-day reduction from an assumed eight-day stay. The published sample-size reassessment retained this target while relying on baseline information and adjusted modelling.¹

Secondary outcomes included hospital mortality, device-associated infections, ventilator-free days and care-process measures. Most secondary analyses included readmissions, producing 7,969 and 8,478 admissions, whereas mortality and the primary analysis used first admissions. Unit performance was assessed using standardised mortality and resource-use measures. Sensitivity analyses examined readiness for discharge, discharge with death as a competing event, and inclusion of readmissions. Secondary comparisons were exploratory and were not adjusted for multiplicity.^{1,3}

During a median 20 months of programme delivery, tele-rounds occurred on 32,074 of 46,920 eligible patient-days, or 68%. Weekends and national holidays were excluded from that denominator. Median consultation duration was eight minutes per patient. Local teams accepted 111,901 of 151,723 recommendations, or 74%. Six units received visits on fewer than 70% of eligible patient-days; five accepted fewer than 70% of recommendations. The report records 149 performance meetings, representing 62% of those expected.¹

Patients were approximately 61 years old on average and most admissions were medical. In the intervention period, invasive ventilation at admission was more frequent in telemedicine units than controls, 48.7% versus 42.2%. Important differences also existed in admission type, cancer prevalence and functional independence. Almost three-quarters of intervention-period patients were enrolled after the pandemic began. The timing of COVID-19 admissions differed between groups, with a greater early burden in telemedicine units and later burden in controls.^{1,3}

Mean critical care stay was 8.1 days with telemedicine and 7.1 days with usual care. The adjusted percentage change was 8.2% (95% CI -5.4% to 23.8%; P=0.24), providing no

evidence of shortening. An additional adjusted median difference was zero days (95% CI -2 to 3). Readiness for discharge and competing-risk analyses were similarly inconclusive; the cause-specific hazard ratio for discharge was 0.97 (95% CI 0.82–1.14).^{1,3}

Hospital mortality was 3,106/7,471 (41.6%) versus 3,119/7,759 (40.2%). The adjusted odds ratio was 0.93 (95% CI 0.78–1.12), with an adjusted absolute difference of -1.2 percentage points (-4.2 to 1.9). No secondary or exploratory outcome showed a statistically convincing benefit. There was also no persuasive treatment interaction in the reported patient, unit or time subgroups. The investigators concluded that this intensivist-led telemedicine programme had not reduced critical care length of stay.^{1,3}

Critique

The strongest feature is the allocation of whole units with concurrent controls. Tele-rounds affect team behaviour and daily organisation, so randomising individual patients within the same unit would invite contamination. A parallel cluster design addresses that problem while protecting more effectively against secular improvements than a simple before-and-after comparison. Broad patient inclusion, a consent waiver and retention of all randomised units further strengthen the evaluation of a service intended to operate across an entire critical care population.¹

The apparent size of the dataset should nevertheless be interpreted correctly. More than 15,000 patients were observed after allocation, but only 30 independent units were randomised. Outcomes within a unit share staffing, resources and clinical practices. The investigators used mixed models, allowed changes over time to vary between clusters and applied a small-sample degrees-of-freedom correction. These are important safeguards; treating every patient as an independent randomised observation would substantially overstate precision. The baseline patients helped characterise units and inform analysis, rather than enlarging the post-allocation comparison by another 1,794 patients.^{1,3}

Restricted randomisation achieved similarity in several unit-level characteristics but did not ensure identical patient populations. Cancer affected 3.7% of intervention-period patients in telemedicine units and 13.2% in controls; emergency surgery and invasive ventilation were more common with telemedicine. Several imbalances were already evident during baseline observation. Adjustment for severity, admission type, ventilation and baseline performance is therefore central to interpretation. The adjusted mortality estimate pointing in a different direction from the crude percentages is not, by itself, an arithmetic mistake or evidence of a survival benefit.^{1,3}

Length of stay is relevant to service capacity but is not an uncomplicated measure of patient benefit. Earlier recovery can shorten it, but so can earlier death, and a patient ready for discharge may remain because another bed is unavailable. Conversely, saving a patient who would otherwise die quickly can lengthen stay. TELESCOPE appropriately supplemented its primary outcome with mortality, readiness for discharge and competing-risk analyses. Their broadly consistent results reduce concern that a major benefit was simply concealed by one discharge convention, although none establishes equivalence.^{1,3}

The primary 8.2% estimate comes from a model of log-transformed stay with adjustment, rather than dividing the two displayed arithmetic means. It should not be reported as an extra 8.2 days, an absolute percentage-point difference, or a proven one-day prolongation. The interval allows a modest proportional shortening and a substantially longer stay. It does not support the expected reduction, but the neutral result cannot establish that every clinically meaningful benefit or harm is absent. The median estimate answers a further, distinct question about the distribution.¹

Intervention exposure was substantial but incomplete. Sixty-eight per cent of eligible weekdays is not 68% of all critical care days. Using the reported totals, 32,074 visits represented approximately 47% of the 67,544 patient-days in the intervention units. This calculation describes the intended service's temporal coverage; it does not redefine weekends as non-adherence. Nor does a missed visit prove that a patient received no specialist input through other contacts. The appropriate conclusion concerns the tested weekday programme, not continuous remote surveillance or round-the-clock specialist staffing.¹

Acceptance is also distinct from implementation. The local team entered whether it accepted the advice, but the trial did not independently establish that each recommendation was delivered. Clinical disagreement may be appropriate, and unavailable staff or equipment can prevent an agreed plan from being carried out. It would therefore be misleading to attribute the neutral result solely to clinicians rejecting advice. Equally, restricting an analysis to units with high acceptance would select units after allocation and could confound treatment effects with motivation, resources and leadership.^{1,3}

The absence of convincing process improvement is informative. For example, the proportion of days in the oxygen saturation target range differed markedly between groups after allocation, but it had differed even more at baseline. The adjusted estimate did not establish a programme effect. Counts of consultations, recommendations or educational-material accesses cannot substitute for a demonstrated change in care.

Some process definitions were also broad: the sedation range extended from a Richmond Agitation-Sedation Scale score of -3 to +1, and the nutrition measure recorded any oral or enteral intake rather than nutritional adequacy.^{1,3}

The pandemic complicates, but does not nullify, the trial. Its timing and impact differed across sites, potentially changing case mix, workload and the ability to attend rounds. Concurrent controls and adjustment for recruitment batches offered protection, and exploratory pre-pandemic and pandemic analyses did not reveal clear heterogeneity. Those subgroup results cannot prove that COVID-19 had no effect on delivery or estimate the treatment effect under an unobserved pandemic-free service. Shared COVID-19 training also narrowed one part of the intervention contrast while retaining the distinction in rounds and feedback.^{1,3}

Generalisation requires similar organisational conditions. The units were public, generally community-based, relatively small and without the established specialist-led service being introduced. Volunteer participation and the large number of unsuccessful initial invitations may further select for local willingness. Results cannot establish the value of adding telemedicine to already well-staffed specialist units, nor whether remote support can replace bedside staff. Transferred patients were counted alive at the end of the observed hospital episode, so the mortality result is not complete 90-day all-cause survival across subsequent facilities.^{1,3}

Finally, hardware supplied without charge is not a demonstration that the programme costs nothing. Remote specialist time, local attendance, technical support and displaced work all have consequences. The report does not establish cost-effectiveness, staff benefit or patient experience. Further evaluation should connect an identified local need with verified changes in care and patient-important outcomes. TELESCOPE shows that delivering consultations at scale is feasible; it also shows that implementation alone does not justify a claim of improved clinical efficiency.^{1,2}

Body of Evidence

Earlier favourable evidence often came from programmes introduced alongside broader changes in care. Lilly and colleagues studied 6,290 adults across seven units at a single academic centre during a sequential telemedicine implementation. Hospital mortality was 13.6% before and 11.8% after implementation, with a much larger adjusted association, odds ratio 0.40 (95% CI 0.31–0.52). Several care processes and hospital stay also improved. The intervention involved electronically supported practice changes and response to alerts. This influential observational evaluation supported further investigation, but its before-and-after structure could not separate the programme from all contemporaneous organisational changes or residual differences in patients.⁴

Nassar and colleagues used concurrent controls when evaluating telemedicine in eight intervention units across seven US Veterans Affairs hospitals. They compared 3,355 patients in intervention units with 3,584 in control units across pre-implementation and post-implementation periods. Neither mortality nor length of stay showed a convincing improvement attributable to telemedicine in adjusted analyses. This was still an observational study, but its control group helped distinguish programme-associated changes from trends occurring elsewhere. The substantially lower mortality of that population also limits direct comparison with the Brazilian units in TELESCOPE. Taken together, these two studies illustrate why the observational literature could generate conflicting impressions.⁵

ERIC supplied randomised evidence from Germany before TELESCOPE. Twelve clusters were allocated to sequences in a stepped-wedge telemedicine trial, although only ten recruited patients. Among 1,463 participants, the programme improved adherence to seven of eight co-primary quality indicators in the principal analysis, using confidence intervals adjusted for the eight comparisons. It did not establish improved adherence to the daily multidisciplinary-round indicator. The analysis followed actual implementation timing, and effects were sensitive to time and local context. Mortality and length of stay were secondary outcomes without convincing improvement. ERIC therefore supports the possibility of changing selected care processes through telemedicine, while leaving the translation from process adherence to patient benefit unresolved.⁶

CHECKLIST-ICU provides a related Brazilian organisational comparison. In 118 units, a programme of daily checklists, goal setting and clinician prompting was tested against usual care. The randomised phase included 6,761 patients; hospital mortality was 32.9% versus 34.8%, with an adjusted odds ratio of 1.02 (95% CI 0.82–1.26). Some exploratory measures of care and team climate improved, but mortality did not. This was not a telemedicine intervention. Its relevance is that even a structured programme directed at

established care processes may fail to change a broad clinical outcome. Neither attractive components nor improved selected indicators guarantee a mortality effect.⁷

The distinction between service delivery, care processes and clinical outcomes links these studies. A programme may successfully increase specialist contact, yet leave treatment unchanged; it may change treatment, yet produce too small an effect to alter length of stay or survival. Conversely, a neutral average effect can coexist with a useful local service, provided that benefit is demonstrated rather than inferred from enthusiasm or technology uptake. TELESCOPE supplies a rigorous estimate for one important model and makes automatic claims of shorter stay difficult to sustain.^{1,2}

As of December 2024, the evidence supported specifying what remote clinicians do, when they are available, how advice is implemented and which local gap the programme addresses. Future trials need a credible link between those features and outcomes, including resource use and consequences for bedside teams. Independently observing implementation would help identify where that link fails, without assuming that clinician agreement is sufficient evidence of changed practice. TELESCOPE does not settle every form of remote critical care, but it sets a clear evidential standard: access to expert advice should not be treated as synonymous with improved patient outcomes.^{1,2}

Do weekday intensivist-led tele-rounds routinely shorten critical care stay?

Not on the evidence from TELESCOPE. Its combined programme did not shorten stay or establish other clinical benefits; any proposed service needs evaluation against its local needs, delivery model and outcomes.

References

1. Pereira AJ, Noritomi DT, dos Santos MC, Corrêa TD, Ferraz LJR, Schettino GPP, et al. Effect of tele-ICU on clinical outcomes of critically ill patients: the TELESCOPE randomized clinical trial. *JAMA* 2024;332(21):1798–1807.
2. Vranas KC, Kahn JM. Evaluating complex technological innovations in critical care—current challenges and future directions. *JAMA* 2024;332(21):1794–1795.
3. Pereira AJ, Noritomi DT, dos Santos MC, Corrêa TD, Ferraz LJR, Schettino GPP, et al. Supplement 3 to: Effect of tele-ICU on clinical outcomes of critically ill patients: the TELESCOPE randomized clinical trial. *JAMA* 2024;332(21):1798–1807.

4. Lilly CM, Cody S, Zhao H, Landry K, Baker SP, McIlwaine J, et al.; University of Massachusetts Memorial Critical Care Operations Group. Hospital mortality, length of stay, and preventable complications among critically ill patients before and after tele-ICU reengineering of critical care processes. *JAMA* 2011;305(21):2175–2183.
5. Nassar BS, Vaughan-Sarrazin MS, Jiang L, Reisinger HS, Bonello R, Cram P. Impact of an intensive care unit telemedicine program on patient outcomes in an integrated health care system. *JAMA Intern Med* 2014;174(7):1160–1167.
6. Spies CD, Paul N, Adrion C, Berger E, Busse R, Kraufmann B, et al.; ERIC Study Group. Effectiveness of an intensive care telehealth programme to improve process quality (ERIC): a multicentre stepped wedge cluster randomised controlled trial. *Intensive Care Med* 2023;49(2):191–204.
7. Writing Group for the CHECKLIST-ICU Investigators and the Brazilian Research in Intensive Care Network (BRICNet). Effect of a quality improvement intervention with daily round checklists, goal setting, and clinician prompting on mortality of critically ill patients: a randomized clinical trial. *JAMA* 2016;315(14):1480–1490.

Possible Reporting Inconsistencies. The nutrition outcome is labelled enteral or parenteral feeding in the tables, but its stated definition is any oral or enteral intake. In Supplementary Table 8, the thromboprophylaxis fractions do not reproduce the displayed percentages: for example, 43,348/67,367 is 64.3%, not 85.3%. The appropriate eligible-day denominators require clarification; the chapter does not infer corrected values.

CYCLE

Kho ME, Berney S, Pastva AM, Kelly L, Reid JC, Burns KEA, et al. Early in-bed cycle ergometry in mechanically ventilated patients. NEJM Evid 2024;3(7):EVIDoa2400137.

Introduction

Survival from critical illness can leave patients with substantial weakness and difficulty performing ordinary activities. Rehabilitation therefore has an objective that mortality alone cannot capture: helping people recover their ability to move and function. Providing useful exercise during invasive ventilation is challenging, however. Sedation limits participation, physiological instability interrupts treatment, and physiotherapy must compete with procedures and other demands on staff. An intervention that can begin before patients are ready to stand has considerable practical appeal. Whether movement delivered at this stage produces meaningful recovery remains an empirical question.¹

In-bed cycle ergometry offers passive, assisted and active movement through the same device. The motor can move a sedated patient's legs; an awake patient can contribute effort without first transferring out of bed. These modes should not be assumed to deliver the same training stimulus. Time attached to a cycle does not establish the amount of active exercise, and additional equipment may change the time available for ordinary physiotherapy. Evaluating the intervention requires a clear account of both its delivered dose and its comparator.^{1,2}

CYCLE tested the addition of early cycling to usual physiotherapy across sixteen intensive care units in three countries. Its primary assessment addressed physical function shortly after discharge from intensive care, rather than strength alone or the speed of achieving a single mobilisation milestone. The trial offers an important test of a feasible rehabilitation strategy in contemporary practice. Interpreting it requires attention to who survived to assessment, how missing measurements were handled, how much cycling was actually active, and whether a short course can reasonably answer questions about longer or more intensive rehabilitation.^{1,2}

Synopsis

CYCLE was a multicentre randomised trial in Canada, Australia and the United States, recruiting between November 2016 and May 2023. Adults had to be within four calendar days of starting invasive ventilation and seven days of intensive care admission, with at least two further intensive care days anticipated. They had to walk independently before admission, with an aid permitted. Important exclusions included conditions preventing safe cycling, substantial pre-existing neurological or cognitive impairment, anticipated very high mortality and persistent physiological instability. Of 6,069 patients screened, 363 were randomised; three assigned to cycling were subsequently excluded as enrolled in error, leaving 178 intervention and 182 control patients. The total includes forty-six participants from an internal pilot.^{1,2}

Allocation used a concealed web-based system, stratified by centre and age. Patients and treating physiotherapists knew the assignment, while outcome assessors were intended to remain unaware. Mean age was approximately sixty-two years, 43% were women, and 89% had medical admissions. Respiratory diagnoses accounted for 44%. Most had lived independently at home, and about 15% were frail before admission. All received invasive ventilation at randomisation; 57% also received inotropes or vasopressors and 12% renal replacement therapy. Eligibility therefore selected for prior functional independence without limiting recruitment to patients needing respiratory support alone.^{1,2}

The intervention comprised thirty minutes of cycling on weekdays, excluding holidays, alongside usual physiotherapy. A motorised device permitted passive, assisted or active movement, with patient effort self-paced and resistance kept constant. Cycling continued until intensive care discharge, a maximum of twenty-eight days, or the ability to march on the spot on two consecutive days. Physiological safety criteria were assessed before treatment, and sessions could be withheld or stopped. Controls received usual physiotherapy without cycling. Ordinary treatment in both groups could include movement exercises, strengthening, transfers, mobilisation and respiratory physiotherapy, according to local practice.^{1,2}

At least one cycling session was delivered to 164 of 178 intervention patients. It began a median two days after ventilation started; recipients received a median three sessions, averaging 27.2 minutes each. Recorded cumulative cycling time was a median ninety minutes. Both groups received a median four usual physiotherapy sessions, but their mean duration was 23.7 minutes with cycling and 29.1 minutes in controls. Among 138 cycling recipients with the necessary mode data, eighty-one contributed active effort at some point. Active cycling occurred on 200 of 687 recorded days. Thus, delivery was early

and usually close to the planned session duration, but most recorded cycling days did not include active effort.^{1,2}

The primary outcome was the Physical Function ICU Test-scored, or PFIT-s, three days after intensive care discharge. Its four components assess arm strength, leg strength, assistance needed to stand and stepping cadence, producing a score from zero to ten, with higher scores indicating better function. The investigators regarded a one-point difference as clinically important. The primary regression adjusted for age and centre and used multiple imputation for missing survivor assessments. The trial recruited 360 eligible patients with the aim of retaining sufficient survivors for its planned functional comparison.^{1,2}

There were 294 survivors at the primary assessment time, 146 and 148 respectively. Complete PFIT-s measurements were available for 111 and 106, leaving seventy-seven missing survivor measurements to be imputed. Observed means were 7.7 and 7.5. The adjusted difference was 0.23 points favouring cycling, with a 95% confidence interval from -0.19 to 0.65 and $P=0.29$. The complete-case estimate was 0.16 points, with a confidence interval from -0.30 to 0.63. Among the 171 assessments completed by blinded assessors, the adjusted difference was 0.39 points, with a confidence interval from -0.12 to 0.91. None established a benefit.^{1,2}

Other assessments did not demonstrate a sustained functional advantage. Median PFIT-s at hospital discharge was 8.8 in both groups. Intensive care acquired weakness at the primary time point occurred in 20 of 106 assessed cycling patients and 20 of 116 assessed controls, with an adjusted odds ratio of 1.14 and a 95% confidence interval from 0.64 to 2.03. An earlier assessment at awakening did yield a nominally positive PFIT-s difference of 0.72 points, with a confidence interval from 0.01 to 1.35. This was one of numerous secondary comparisons, rather than the primary result. Ninety-day patient-reported outcomes were collected only in the later recruitment cohort and did not establish a lasting benefit.^{1,2}

Median ventilation duration was five versus 5.5 days, intensive care stay eleven days in both groups, and hospital stay 21.5 versus twenty-one days. Hospital mortality was 39/178, or 21.9%, with cycling and 40/182, or 22.0%, with usual physiotherapy; the hazard ratio was 0.92, with a 95% confidence interval from 0.59 to 1.44. These results did not demonstrate a survival advantage. Among assessed hospital survivors, the proportion discharged to the same or a better residential setting was approximately 60% in each group.^{1,2}

No serious adverse events attributed to rehabilitation were reported. Such events were defined as cardiac arrest, unplanned extubation or a fall to the knees. Other attributed adverse events occurred in 16/174 intervention patients who received rehabilitation and 10/171 treated controls, or 9.2% versus 5.8%. Eleven cycling recipients experienced an event during cycling. Sessions were supervised, and physiotherapists could stop treatment for symptoms or physiological changes. The investigators concluded that adding this early cycling programme did not improve physical function three days after intensive care discharge.^{1,2}

Critique

CYCLE provides substantially stronger evidence than a simple demonstration that ventilated patients can use a cycle. Concealed randomisation, recruitment across sixteen units and a functional primary outcome support a clinically useful comparison. The investigators documented ordinary physiotherapy as well as cycling, allowing the actual treatment contrast to be assessed. The absence of control-group cycling also limits direct contamination. These strengths make the trial informative about offering this programme to eligible patients in units already providing rehabilitation.^{1,2}

The primary estimand nevertheless concerns function among survivors. Death prevents the functional assessment and is not simply another missing questionnaire. Similar mortality between groups is reassuring, but it cannot guarantee that the survivors compared have equivalent underlying prognoses. An intervention could, in principle, change who survives without producing a clear mortality difference in a study this size. The primary estimate should therefore be described as a survivor functional comparison, rather than an overall recovery benefit incorporating death.^{1,2}

Missing assessments introduce a separate limitation. Seventy-seven of 294 survivors, approximately 26%, lacked a complete primary measurement. Multiple imputation used available baseline and functional information, which is preferable to treating these patients as if they had never been enrolled. Its validity still depends on how adequately observed variables explain missingness. Patients discharged quickly and patients too unwell to complete testing can have very different outcomes. Agreement with complete-case analysis is helpful, but both approaches can miss a bias related to unmeasured reasons for non-assessment.²

The sensitivity analysis assigning a zero score to patients who died addresses mortality only partly. Its published denominators total 283, corresponding to the 217 complete assessments plus sixty-six deaths, rather than all 360 eligible randomised patients. It therefore does not recover the missing survivor measurements. Moreover, the data did

not meet the stated linear regression assumptions, and the table reports medians without the primary-style adjusted effect estimate. It should not be presented as a definitive combined survival-and-function analysis that resolves every concern about the survivor outcome.²

Blinding was also incomplete. Although independent assessment was intended, only 171 of the 217 complete primary assessments were blinded. Functional testing incorporates standardised tasks, but encouragement, judgement about assistance and patient effort can still matter. Restricting analysis to blinded assessments gave a compatible, uncertain estimate. That finding supports the overall interpretation without recreating the original randomised comparison: the subset was determined after allocation and excluded more than half the eligible population.^{1,2}

The confidence interval is more informative than the non-significant P value alone. Under the primary analysis assumptions, its upper limit of 0.65 points excludes the one-point mean benefit used to motivate the study. CYCLE therefore weighs against a benefit of that magnitude from this delivered programme at this assessment time. It does not prove identical outcomes, exclude smaller average benefits, or establish that particular individuals cannot benefit. Nor does it provide equally precise answers about mortality, long-term disability or adverse events, for which the sample and event counts were much smaller than would be needed to exclude modest differences.^{1,2}

Intervention dose requires careful interpretation. Three sessions and ninety cumulative minutes describe a short exposure, and much of the movement was passive. The eighty-one patients who ever contributed active effort form a selected subset; their performance cannot establish the effect of assigning active cycling. A longer course, progressive resistance or greater patient-generated work would constitute different interventions. They remain reasonable research questions, but their possible efficacy cannot be used to convert this neutral trial into evidence that cycling works when delivered properly.^{1,2}

The comparator also matters. Control patients began physiotherapy early and received similar numbers of sessions. Ordinary session duration was shorter when cycling was added, so the trial did not simply add a fixed thirty minutes to otherwise identical therapy. It tested the net effect of incorporating cycling into a working rehabilitation service. Setup and pack-down required additional time, averaging about nineteen minutes per cycling session. Staffing availability was the leading recorded reason why eligible patients were not randomised. These details make opportunity cost a practical consideration, although the trial does not provide a complete cost-effectiveness estimate.^{1,2}

Treatment fidelity was affected by both medical exemptions and missed opportunities. The planned combined cycling-and-physiotherapy programme was delivered on 583 of 1,192 planned days; 299 were temporarily exempt and 310 missed. The latter represents 26% of all planned days, rather than 26% of days remaining after medical exemptions. An adherence sensitivity analysis also counted temporary exemptions as compliant. Its neutral result cannot identify an effective biological dose or establish that receiving more therapy would have been ineffective; adherence is influenced by recovery, illness severity and service capacity after randomisation.²

Secondary and subgroup findings merit restraint in both directions. The awakening PFIT-s result means that every functional comparison was not simply null, but multiplicity, missing measurements and variable assessment timing make it exploratory. The larger estimated effect among frail participants was imprecise and based on few complete assessments. It does not identify a subgroup for routine treatment. Ninety-day assessments were unavailable by design for the first forty-six participants, and further measurements were missing in the later cohort, limiting claims about durability.^{1,2}

Finally, the safety findings apply to screened patients receiving monitored sessions with explicit stopping criteria. Zero serious attributed events is reassuring for feasibility, but does not demonstrate safety in patients excluded because of instability or inability to cycle. Event attribution was unblinded, and the non-serious event proportions were numerically higher with the intervention. The practical conclusion is that this supervised programme was deliverable without a detected serious therapy-related event and without the expected functional gain. It provides no basis for abandoning individualised mobilisation, or for routinely adding a device in expectation of improved recovery.^{1,2}

Body of Evidence

An early trial by Burtin and colleagues randomised ninety patients to usual physiotherapy with or without daily cycling. Recruitment occurred after cardiorespiratory stability, from the fifth intensive care day, with at least a further week in intensive care anticipated. The intervention comprised twenty minutes of passive or active cycling. Strength and function at intensive care discharge were not clearly different, but hospital-discharge walking distance, quadriceps strength and reported physical function favoured cycling among survivors. This was encouraging, small-scale evidence in a population selected for a longer stay. It does not establish that starting earlier and delivering a median three sessions, as in CYCLE, should produce the same result. Conversely, CYCLE's broader early recruitment does not precisely reproduce this later, selected treatment strategy.³

Schweickert and colleagues studied early physical and occupational therapy in 104 ventilated patients. Both groups received daily interruption of sedation; the intervention added coordinated early therapy, while controls received therapy when ordered by the clinical team. Independent function at hospital discharge was achieved by 29/49 intervention patients and 19/55 controls, or 59% versus 35%, with an odds ratio of 2.7 and a 95% confidence interval from 1.2 to 6.1. The study supported integrating rehabilitation with an approach that enabled participation. It did not isolate cycle ergometry, and its usual-care contrast differs from CYCLE's early physiotherapy in both arms. Evidence for that combined service cannot be transferred directly to every additional exercise device.⁴

TEAM subsequently randomised 750 ventilated adults to increased early active mobilisation or usual care. Daily active mobilisation averaged 20.8 versus 8.8 minutes, and intervention patients stood earlier, but days alive and out of hospital at day 180 were not improved: medians were 143 and 145, with an estimated difference of –2 days and a 95% confidence interval from –10 to 6. Potentially mobilisation-related adverse events occurred in 9.2% versus 4.1%, and serious adverse events in seven versus one patient. TEAM cautions against equating more activity or earlier achievement of a milestone with better overall recovery. Its active mobilisation programme differed from CYCLE's largely passive, self-paced cycling. The trials jointly support testing rehabilitation dose and patient benefit rather than assuming that greater intensity is necessarily better.⁵

The updated cycling review published in October 2024 included thirty-three randomised trials and 3,274 patients. Pooled physical function favoured cycling at intensive care discharge, with a standardised mean difference of 0.33 and a 95% confidence interval from 0.05 to 0.62, and after hospital discharge, with a difference of 0.23 and a confidence interval from 0.04 to 0.42. Certainty was low for both. Hospital-discharge function remained uncertain. Hospital stay was an estimated 1.48 days shorter, with a confidence interval from 0.47 to 2.49 days shorter and moderate certainty; most other outcomes had low or very low certainty. The analysis included CYCLE itself, so its pooled result is not independent confirmation alongside CYCLE.⁶

The review combined different populations, cycling modes, co-interventions and functional instruments. A standardised mean difference expresses an effect relative to variation in the measurements; it is not a change in PFIT-s points and cannot be compared directly with CYCLE's one-point threshold. Intensive care discharge also differs from assessment three days later. Many trials were small, and some delivered cycling within a wider rehabilitation package. Separately pooled adverse-event proportions for cycling and control sessions came from different numbers of trials and should not be

treated as a randomised relative-risk comparison. Taken together, the evidence leaves room for a modest benefit in some settings but does not establish routine early cycling as an effective addition to already available physiotherapy. Future trials need to distinguish active work from passive movement and demonstrate an improvement that patients can sustain beyond the treatment session.^{1,6}

Should early in-bed cycling routinely be added to physiotherapy during invasive ventilation?

CYCLE does not support routine addition to improve early post-intensive-care physical function. Continue individualised rehabilitation; any cycling programme should account for patient tolerance, active exercise dose and staff time.

References

1. Kho ME, Berney S, Pastva AM, Kelly L, Reid JC, Burns KEA, et al. Early in-bed cycle ergometry in mechanically ventilated patients. *NEJM Evid* 2024;3(7):EVIDoA2400137.
2. Kho ME, Berney S, Pastva AM, Kelly L, Reid JC, Burns KEA, et al. Supplementary appendix to: Early in-bed cycle ergometry in mechanically ventilated patients. *NEJM Evid* 2024;3(7):EVIDoA2400137.
3. Burtin C, Clerckx B, Robbeets C, Ferdinande P, Langer D, Troosters T, et al. Early exercise in critically ill patients enhances short-term functional recovery. *Crit Care Med* 2009;37(9):2499–2505.
4. Schweickert WD, Pohlman MC, Pohlman AS, Nigos C, Pawlik AJ, Esbrook CL, et al. Early physical and occupational therapy in mechanically ventilated, critically ill patients: a randomised controlled trial. *Lancet* 2009;373(9678):1874–1882.
5. TEAM Study Investigators and the ANZICS Clinical Trials Group. Early active mobilization during mechanical ventilation in the ICU. *N Engl J Med* 2022;387(19):1747–1758.
6. O’Grady HK, Hasan H, Rochweg B, Cook DJ, Takaoka A, Utgikar R, et al. Leg cycle ergometry in critically ill patients—an updated systematic review and meta-analysis. *NEJM Evid* 2024;3(12):EVIDoA2400194.

Possible Reporting Inconsistencies. The main results text gives the primary adjusted difference as –0.23; the abstract, Table 3 and supplement give +0.23, used here. Table 4 lists 918 control physiotherapy days, versus 979 in Table 2 and Supplementary Table S13; patient-level event proportions are used here. A main-table footnote dates ninety-day outcomes from intensive care discharge; the heading and supplement specify randomisation.

Summarising, critiquing and putting into context the best critical care trials of 2024



Rob Mac Sweeney