



Critical Care Reviews Book 2024

The Best Critical Care Trials of 2023
First Edition





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 2024 brings together a selection of the major critical care trials published in 2023. 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 2023. 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 2023. 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
EARLYDRAIN	2
OXY-TC	12
TENSION	22
VASE	31
DASH	40
■ Circulatory Trials	49
MINT	50
ECLS-SHOCK	59
■ Resuscitation Trials	68
TAME	69
STEROHCA	78
ARREST	87
INCEPTION	96
■ Airway Trials	105
NICO	106
DEVICE	115
AMBROISIE	124
REMICRUSH	133
■ Respiratory Trials	143
AMIKINHAL	144
SiVENT	153
EXTUB-OBESE	162
CAPE-COD	171
ICONIC	180
■ Gastrointestinal Trials	189
EPIPAN	190
■ Nutrition Trials	199
TGC-FAST	200
NUTRIREA-3	209
EFFORT-Protein	218

Table of Contents

■ Liver Trials	227
AntibioCor	228
■ Haematology Trials	237
PACER	238
STRATUS	246
■ Sepsis Trials	255
REMAP-CAP Simvastatin	256
LOVIT-COVID	265
CONFIDENT	274
STRESS-L	283
ACORN	292
MERCY	301
CLOVERS	310
ACTIV-1-IM	319
Mupirocin–Iodophor Swap	329
■ Trauma Trials	338
SHINE-TRAUMA	339
CRYOSTAT-2	348
UK-REBOA	357
RESTRIC	366
PATCH-TRAUMA	375
PROCOAG	384
TOP-ART	393
■ Paediatric Trials	403
OXY-PICU	404
SafeBoosC-III	413
PARIS-2	422
■ Miscellaneous Trials	431
Early Mobilisation in ICU	432
PICS-H	441



Neuro Trials

EARLYDRAIN

Wolf S, Mielke D, Barner C, Malinova V, Kerz T, Wostrack M, et al; EARLYDRAIN Study Group. Effectiveness of Lumbar Cerebrospinal Fluid Drain Among Patients With Aneurysmal Subarachnoid Hemorrhage: A Randomized Clinical Trial. JAMA Neurol 2023;80(8):833–842.

Introduction

Securing a ruptured intracranial aneurysm addresses rebleeding, but leaves blood within the subarachnoid space. Hydrocephalus, intracranial hypertension and delayed cerebral injury may still cause death or disability despite successful aneurysm occlusion. Limiting these secondary insults might therefore improve recovery. The challenge is to identify an intervention that changes neurological outcome rather than merely a physiological or radiological measurement.^{1,2}

Cerebral vasospasm has traditionally occupied a central position in this process. However, arterial narrowing, clinical deterioration and cerebral infarction are related but distinct phenomena. The consensus approach to research outcomes emphasises infarction and functional recovery, while interpreting angiographic or transcranial Doppler evidence of vasospasm alongside these patient-relevant measures. This distinction matters when evaluating a treatment intended to remove blood and its degradation products: demonstrating less vasospasm would not, by itself, establish clinical benefit.³

Lumbar cerebrospinal fluid drainage offers a plausible means of accelerating blood clearance from the basal subarachnoid spaces. The proposed mechanism differs from simply draining more cerebrospinal fluid: dependent drainage might remove sedimented blood more effectively than diversion through an external ventricular drain. Earlier observational findings were encouraging, but could not establish that the drainage strategy itself caused better outcomes. The subsequent LUMAS randomised trial did not establish an improvement in six-month functional outcome.^{1,4,5}

The decision is also constrained by safety. Lumbar drainage in a patient with impaired communication between intracranial and spinal cerebrospinal fluid compartments may create a hazardous pressure gradient. An effective strategy must therefore combine appropriate patient selection, timing after aneurysm occlusion and controlled drainage. EARLYDRAIN tested whether adding early lumbar drainage to standard treatment

improved six-month neurological outcome across a broader range of clinical severity, rather than selecting vasospasm alone as its primary endpoint.^{1,2}

Synopsis

EARLYDRAIN was an investigator-initiated, pragmatic, multicentre, parallel-group, open-label randomised trial with blinded outcome assessment. Nineteen referral hospitals in Germany, Switzerland and Canada enrolled patients between 31 January 2011 and 24 January 2016. Participating centres provided continuous neurosurgical and neurocritical care and performed at least 30 aneurysm procedures annually. Follow-up finished in July 2016; queries concerning missing case-report-form information continued until September 2020. Support came from the Center for Stroke Research and unrestricted Orsan Medical Systems funds, with staff supported by their institutions. Funders had no role in design, conduct, analysis or publication.²

Adults with a first, computed tomography (CT)-visible aneurysmal subarachnoid haemorrhage were eligible irrespective of clinical grade, provided their premorbid modified Rankin Scale (mRS) score was 0 or 1 and the aneurysm could be treated within 48 hours of haemorrhage. Important exclusions included absent or compressed basal cisterns, therapeutic anticoagulation or another contraindication to lumbar drainage, non-aneurysmal haemorrhage, pregnancy, participation in another interventional trial and serious competing disease or limited life expectancy. Consent was obtained from patients or authorised representatives, with an independent-physician and subsequent confirmation pathway where permitted.^{2,6}

An internet-based system allocated patients 1:1 using permuted blocks of six, with the block size undisclosed to investigators and no stratification. The intervention added a lumbar drain after clipping or coiling. Drainage began only after a postprocedural CT indicated that it was safe, and was initiated within 72 hours of the haemorrhage. The final report recommended 5 mL/h during the first eight days. Both groups could receive an external ventricular drain (EVD), with its drainage volume determined by clinical need. Controls did not receive routine prophylactic lumbar drainage, although clinically indicated rescue drainage was permitted.²

When ventricular and lumbar pressure were measured together, both systems were zeroed at the external auditory canal. Lumbar diversion was to be deferred if intracranial pressure (ICP) exceeded 20 mmHg or the pressure difference between compartments exceeded 5 mmHg. Neurological deterioration potentially related to drainage required immediate interruption and reassessment. Other treatment followed local protocols. Transcranial Doppler surveillance and vascular imaging for suspected vasospasm, or

around days 7–10, informed discretionary rescue treatment with intra-arterial vasodilators or angioplasty.^{2,6}

The primary endpoint was an unfavourable outcome at six months, defined as mRS 3–6 rather than 0–2; death scores 6. A local investigator uninvolved in acute care and blinded to the clinical course assessed outcome by interview or personal visit. The main secondary endpoint was a new infarction on the final brain scan before acute-care discharge, absent from the scan after aneurysm treatment. Radiologists were masked to allocation. Other outcomes included mortality, additional functional scales, clinically suspected and instrumentally detected vasospasm, permanent shunting and infections.²

The planned sample of 300 provided 85% power at a two-sided 5% significance level to detect a reduction in unfavourable outcome from 50% to 33%, an absolute difference of 17 percentage points. A safety assessment after 150 participants was planned. There was no separate statistical analysis plan; analytical methods were described within the protocol. Primary analyses used logistic regression, with odds ratios transformed into risk ratios. Adjustment included age, Hunt–Hess grade and intracerebral or intraventricular haemorrhage. Missing values were not imputed. No multiplicity adjustment was applied, and analyses beyond the primary endpoint were considered exploratory.^{2,6}

There were 307 randomisation records, including duplicate allocations. Twenty were deemed invalid, leaving 144 lumbar-drain and 143 control patients for the analysis described as intention-to-treat. Reasons included consent problems, late aneurysm treatment, unexpected antithrombotic requirements, duplicate or erroneous randomisation and unavailable source data. After 300 allocations, seven known failures had prompted seven additional allocations; recruitment was not stopped for demonstrated efficacy. Complete screening logs were available from only five centres: 591 patients were seen and 219 randomised. Primary outcome assessment was complete for all 287 patients retained in the analysis.^{2,7}

The median age was 55 years and 68.6% were women. The lumbar-drain group contained more patients with favourable clinical grades: Hunt–Hess grades 1–2 occurred in 48.6% versus 37.1%. Conversely, intracerebral haemorrhage was present in 38.9% versus 35.0%, and intraventricular blood in 62.5% versus 59.4%. Approximately half underwent clipping and half coiling. Nimodipine was administered to 98.6% and 99.3%, respectively; magnesium and statins were also used in both groups.^{2,7}

A lumbar catheter was inserted in 131/144 intervention patients (91.0%), but only 109/144 (75.7%) met the per-protocol exposure threshold of at least 480 mL drained

during the first eight days. Nine controls received a lumbar catheter; two reached that threshold. Lumbar drainage began a median of two days after haemorrhage, with a median daily volume of 108 mL. EVDs were used in 70.8% versus 76.9%. Median daily ventricular drainage was lower with the lumbar strategy, 98 versus 171 mL, while total cerebrospinal fluid drainage over eight days was similar, 1170 versus 1202 mL. In patients with available measurements, morning ICP and daily peak ICP averaged 1.5 and 2.9 mmHg lower, respectively.^{2,7}

At six months, the primary endpoint occurred in 47/144 patients (32.6%) assigned lumbar drainage and 64/143 (44.8%) assigned standard care: risk ratio 0.73 (95% CI, 0.51–0.98; P = 0.04). The reported absolute difference was –12 percentage points (95% CI, –23 to –1). Adjustment for baseline severity produced a risk ratio of 0.76 (95% CI, 0.54–1.00; P = 0.047). These estimates are from the main results table.²

New infarction before discharge occurred in 41/144 (28.5%) versus 57/143 (39.9%): risk ratio 0.71 (95% CI, 0.49–0.99; P = 0.04). Six-month mortality was 19/144 (13.2%) versus 25/143 (17.5%): risk ratio 0.75 (95% CI, 0.42–1.28; P = 0.31). Clinically suspected vasospasm occurred in 28.5% versus 33.6%; Doppler-defined vasospasm in 26.9% versus 24.8% of assessed patients; and angiographic vasospasm in 46.0% versus 44.0%. None differed statistically. Permanent shunting by six months was also similar, 28.5% versus 29.4%.²

An unfavourable Glasgow Outcome Scale–Extended score at six months occurred in 26.4% versus 37.8%: risk ratio 0.70 (95% CI, 0.47–0.98). Among surviving patients with measurements, a Barthel Index of 80 or less occurred in 20/125 (16.0%) versus 35/116 (30.2%): risk ratio 0.53 (95% CI, 0.30–0.86). The per-protocol primary-outcome analysis favoured lumbar drainage, 28.4% versus 44.0%, as did the as-treated analysis, 29.7% versus 44.3%. Subgroup estimates did not establish convincing treatment-effect heterogeneity.²

One patient developed a pressure gradient preventing continued drainage, one required surgery for a lumbar insertion-site infection, and one required removal of a retained catheter fragment. No death was attributed to lumbar-drain complications. Suspected infection of any cause occurred in 56/144 (38.9%) versus 52/143 (36.4%): risk ratio 1.07 (95% CI, 0.78–1.39). This endpoint was not a specific measure of drain-associated meningitis.^{2,7}

Critique

EARLYDRAIN provides evidence of benefit on a prespecified, patient-centred endpoint, rather than asking clinicians to infer benefit from improved vasospasm or ICP measurements alone. Central allocation, blinded six-month assessment and complete primary-outcome ascertainment within the analysed population are important strengths. However, the result supports a particular strategy: early lumbar drainage after aneurysm occlusion, in selected patients, with active surveillance and interruption when unsafe. It does not establish that every patient with subarachnoid haemorrhage should receive a lumbar drain.^{2,6}

The comparator is fundamental to interpretation. Most patients in both groups had ventricular drains, and total cerebrospinal fluid output was similar. The trial therefore did not simply compare drainage with no drainage, nor did it test replacing EVDs with lumbar catheters. Adding a lumbar route changed the distribution of drainage between compartments and reduced ventricular output. This is consistent with the proposed importance of where fluid is removed, rather than its total volume. Nevertheless, the pragmatic comparison cannot isolate blood clearance, pressure control or another consequence of the combined management strategy as the cause of better recovery.^{2,7}

The primary point estimate is clinically substantial, but its precision is modest. The absolute confidence interval extends from approximately one to 23 fewer unfavourable outcomes per 100 treated patients. The adjusted result remains just within the conventional significance threshold, with its upper confidence limit rounding to 1.00. These findings should not be dismissed because the P value is close to 0.05, but neither should the point estimate be treated as a reliable promise of benefit to one in every eight patients. The trial was designed around a large 17-percentage-point effect and does not precisely determine the magnitude achievable elsewhere.²

The analysis population requires closer attention than the label intention-to-treat suggests. Twenty randomisation records were removed. Duplicated records and absent consent are not equivalent to excluding an eligible participant because treatment became difficult. Supplementary descriptions also include one patient in each group who died before receiving the intended intervention and whose data were excluded. Removal after allocation can weaken the protection randomisation provides, depending on the reason and its relationship to prognosis or treatment. The available reporting does not establish the direction or magnitude of any resulting bias. Equally, complete follow-up in 287 analysed patients must not be confused with complete inclusion of every allocation.^{2,7}

Incomplete treatment delivery creates a separate problem. Almost one-quarter of the lumbar-drain group did not reach the 480-mL threshold, because of technical barriers, clotting, clinical concerns or other reasons. Some interruptions were appropriate responses to safety conditions built into the strategy. The randomised comparison therefore estimates the effect of offering this monitored approach, including situations in which drainage cannot be delivered. Although per-protocol and as-treated estimates were larger, they do not recover a cleaner randomised treatment effect. Classification by cumulative output selects patients according to events occurring after allocation; ability to continue drainage may itself reflect neurological stability, complications or survival.^{2,7}

Baseline imbalances were not uniformly favourable to either group. Clinical grades tended to favour lumbar drainage, while parenchymal and ventricular blood were somewhat more frequent in that group. Adjustment is consequently useful, but it does not make these issues disappear. The protocol specified regression adjustment in broad terms, rather than providing a separate, detailed analysis plan. The final analyses considered alternative predictors and compared models using the Akaike information criterion. This flexibility does not demonstrate selective analysis, but it makes the prespecified unadjusted comparison especially important and limits how strongly an adjusted P value can be invoked as independent confirmation.^{2,6,7}

Blinded outcome assessment reduces detection bias, but cannot remove differences in care arising from an open-label intervention. Rescue treatment, ventricular drainage and other management remained discretionary. Recorded blood pressure, fluid balance and fever measurements did not reveal important group differences, yet these observations cannot exclude differences in unmeasured care. Conversely, avoiding rigid protocols for every co-intervention makes the study more representative of clinical delivery. The appropriate inference concerns the whole assigned care strategy, not an experimentally isolated effect of a lumbar catheter.^{2,7}

The secondary findings are supportive rather than independent proof. Fewer infarctions and better additional functional measures fit the primary result, but testing was not multiplicity-adjusted and the disability scales assess overlapping aspects of recovery. The Barthel analysis additionally excluded deaths and had two missing survivor measurements. It therefore answers a different question from an endpoint incorporating death. The mortality confidence interval includes both benefit and harm, so a survival advantage remains unestablished. Likewise, directionally similar subgroup estimates do not show that benefit has been demonstrated separately in every severity grade, aneurysm location or treatment subgroup.²



Mechanistically, fewer infarctions despite similar detected vasospasm challenge a simple large-artery explanation. Lower ICP offers another possibility, but morning ICP data were unavailable or unmeasured in 100 patients, and no mediation analysis establishes pressure reduction as the route to improved function. Blood clearance was not quantified sufficiently to confirm the proposed mechanism. Infarction assessment also depended predominantly on CT at discharge rather than uniform magnetic resonance imaging at a fixed time. It is reasonable to regard the imaging findings as biological support, but not to relabel the result as proof that lumbar drainage prevents every form of delayed cerebral ischaemia.^{2,7}

Safety is inseparable from selection and monitoring. Patients with important anatomical contraindications were excluded, and drainage was interrupted for pressure abnormalities or neurological concern. An absence of attributed fatal complications in this modest sample cannot exclude uncommon severe harm outside these conditions. Moreover, the broadly defined infection endpoint cannot establish equivalent rates of device-associated meningitis. The supplementary appendix explains that several common medical complications were not recorded separately as adverse events, and some complications were classified inconsistently between investigators. This limits comprehensive safety comparison even though the specifically described catheter complications were few.^{2,7}

Finally, multicentre participation should not obscure concentration of experience. More than half the patients came from two centres, screening records were incomplete, and poor clinical grade was a reason for non-enrolment despite formal inclusion of all grades. The selected population may therefore differ from consecutive admissions. Almost equal use of clipping and coiling, and recruitment ending seven years before publication, also warrant attention when applying the result. Overall, EARLYDRAIN materially strengthens the case for early adjunctive lumbar drainage, while leaving uncertainty about effect size, optimal candidates and reproducibility of safe delivery.^{2,7}

Body of Evidence

The early observational signal was unusually large. Klimo and colleagues examined a non-randomised cohort of 167 patients, of whom 81 received lumbar drainage and 86 received an EVD alone or no cerebrospinal fluid drainage. Clinical vasospasm occurred in 17% versus 51%, vasospastic infarction in 7% versus 27%, and angioplasty was required in 17% versus 45%. A Glasgow Outcome Scale score of 5 at one to three months after discharge occurred in 71% versus 35%. These findings supplied a rationale for randomised evaluation, but patient selection and other differences in management

could contribute to the apparent benefit. The magnitude of an observational association should not be assumed to represent the effect that a subsequent trial must reproduce.⁴

A small physiological study highlighted why prophylactic drainage cannot be separated from its safety conditions. Staykov and colleagues described 22 patients receiving early lumbar drainage after aneurysm treatment. Three developed a progressive lumboventricular pressure gradient; two showed clinical signs of herniation, which resolved after clamping. The gradient developed at least 12 hours before clinical symptoms. This was not a randomised efficacy comparison and cannot quantify the risk under the EARLYDRAIN protocol. It nevertheless supplies direct clinical support for monitoring communication between compartments and responding to a developing gradient rather than assuming that a technically successful catheter insertion establishes safety.⁸

LUMAS provided the most directly relevant earlier randomised comparison. This single-centre trial allocated 210 patients with World Federation of Neurosurgical Societies grades 1–3 to standard care with or without lumbar drainage. Its primary outcome, delayed ischaemic neurological deficit, was less frequent with drainage: 21.0% versus 35.2% ($P = 0.021$). Poor functional outcome, defined as mRS 4–6, was also less frequent at day 10, 44.8% versus 62.5%, but not at six months, 19.8% versus 18.6% ($P = 0.83$). Thus, LUMAS was positive for its primary endpoint, but did not establish sustained functional benefit. Its milder clinical population and different mRS threshold distinguish it from EARLYDRAIN. Comparing the raw six-month poor-outcome percentages between trials without accounting for that threshold would be misleading.⁵

A 2022 systematic review and meta-analysis by Hulou and colleagues included 17 studies of lumbar drainage. Compared with no drainage, pooled analyses favoured lumbar drainage for early favourable mRS, mortality, clinical vasospasm and ischaemic stroke or delayed neurological deficit. The functional endpoint emphasised the period within one month of discharge, not uniformly six months. This synthesis preceded publication of EARLYDRAIN and did not remove the interpretative limitations of the component designs and drainage comparisons. A pooled favourable result can support the rationale for investigation, but cannot substitute for a well-defined randomised comparison with durable functional outcome as its primary endpoint.⁹

The broader vasospasm literature explains why that endpoint choice matters. CONSCIOUS-2 randomised 1157 surgically clipped patients to clazosentan 5 mg/h or placebo. Despite the drug's earlier evidence of reducing angiographic vasospasm, the six-week composite of death or vasospasm-related morbidity occurred in 21% versus 25% of treated patients, without a statistically significant reduction. Poor functional

outcome at 12 weeks was 29% versus 25%, and mortality was 6% in both groups. Pulmonary complications, anaemia and hypotension were more frequent with clazosentan. This does not provide evidence against lumbar drainage: it tests another intervention. It does, however, illustrate why a plausible vascular mechanism and a favourable surrogate cannot replace demonstration of patient benefit.¹⁰

Taken together, these studies leave a more specific conclusion than either “drainage works” or “vasospasm prevention has failed”. Observational studies established a promising signal, physiological work identified a preventable mode of harm, and LUMAS showed that an early neurological effect need not translate into a six-month advantage. EARLYDRAIN adds randomised evidence of improved six-month outcome in an anatomically selected, mixed-severity population. The unresolved questions concern the size and consistency of that benefit, the patients most likely to benefit, and whether comparable outcomes can be achieved while preserving the selection and monitoring that made the trial strategy feasible.^{1,2}

Should we add early lumbar drainage after aneurysm occlusion in adults with aneurysmal subarachnoid haemorrhage?

Consider this in selected patients, following the trial's safety precautions. EARLYDRAIN improved six-month neurological outcomes but did not establish a survival benefit.

References

1. Bardutzky J, Witsch J, Jüttler E, Schwab S, Vajkoczy P, Wolf S. EARLYDRAIN—outcome after early lumbar CSF-drainage in aneurysmal subarachnoid hemorrhage: study protocol for a randomized controlled trial. *Trials* 2011;12:203.
2. Wolf S, Mielke D, Barner C, Malinova V, Kerz T, Wostrack M, et al; EARLYDRAIN Study Group. Effectiveness of Lumbar Cerebrospinal Fluid Drain Among Patients With Aneurysmal Subarachnoid Hemorrhage: A Randomized Clinical Trial. *JAMA Neurol* 2023;80(8):833–842.
3. Vergouwen MDI, Vermeulen M, van Gijn J, Rinkel GJE, Wijdicks EFM, Muizelaar JP, et al. Definition of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage as an outcome event in clinical trials and observational studies: proposal of a multidisciplinary research group. *Stroke* 2010;41(10):2391–2395.

4. Klimo P Jr, Kestle JRW, MacDonald JD, Schmidt RH. Marked reduction of cerebral vasospasm with lumbar drainage of cerebrospinal fluid after subarachnoid hemorrhage. *J Neurosurg* 2004;100(2):215–224.
5. Al-Tamimi YZ, Bhargava D, Feltbower RG, Hall G, Goddard AJP, Quinn AC, et al. Lumbar drainage of cerebrospinal fluid after aneurysmal subarachnoid hemorrhage: a prospective, randomized, controlled trial (LUMAS). *Stroke* 2012;43(3):677–682.
6. Wolf S, Mielke D, Barner C, Malinova V, Kerz T, Wostrack M, et al; EARLYDRAIN Study Group. Supplement 1: EARLYDRAIN Trial Protocol, protocol versions, statistical analysis description and case report form. Supplement to: *JAMA Neurol* 2023;80(8):833–842.
7. Wolf S, Mielke D, Barner C, Malinova V, Kerz T, Wostrack M, et al; EARLYDRAIN Study Group. Supplement 2: Supplemental Online Content, additional analyses, tables and figures. Supplement to: *JAMA Neurol* 2023;80(8):833–842.
8. Staykov D, Speck V, Volbers B, Wagner I, Saake M, Doerfler A, et al. Early recognition of lumbar overdrainage by lumboventricular pressure gradient. *Neurosurgery* 2011;68(5):1187–1191.
9. Hulou MM, Essibayi MA, Benet A, Lawton MT. Lumbar Drainage After Aneurysmal Subarachnoid Hemorrhage: A Systematic Review and Meta-Analysis. *World Neurosurg* 2022;166:261–267.e9.
10. Macdonald RL, Higashida RT, Keller E, Mayer SA, Molyneux A, Raabe A, et al. Clazosentan, an endothelin receptor antagonist, in patients with aneurysmal subarachnoid haemorrhage undergoing surgical clipping: a randomised, double-blind, placebo-controlled phase 3 trial (CONSCIOUS-2). *Lancet Neurol* 2011;10(7):618–625.

OXY-TC

Payen JF, Launey Y, Chabanne R, Gay S, Francony G, Gergele L, et al; OXY-TC trial collaborators. Intracranial pressure monitoring with and without brain tissue oxygen pressure monitoring for severe traumatic brain injury in France (OXY-TC): an open-label, randomised controlled superiority trial. Lancet Neurol 2023;22(11):1005–1014.

Introduction

Preventing secondary injury is central to the intensive care management of severe traumatic brain injury (TBI). Intracranial pressure (ICP) monitoring identifies intracranial hypertension and allows calculation of cerebral perfusion pressure (CPP), informing treatment intended to preserve brain perfusion. However, acceptable pressure measurements do not necessarily imply adequate tissue oxygenation. Brain hypoxia can occur despite satisfactory ICP and CPP, through systemic oxygen-delivery problems or local disturbances in the injured microcirculation.¹

An intraparenchymal brain tissue oxygen pressure (PbtO_2) probe provides an additional measurement with which to guide treatment. Low values might prompt changes in arterial oxygenation, ventilation, perfusion pressure, temperature or haemoglobin concentration. The attraction is the possibility of identifying a treatable insult that pressure monitoring misses. The difficulty is that a focal oxygen measurement is neither a whole-brain assessment nor a direct demonstration that cellular metabolism has recovered. Raising the measured value and improving neurological outcome are different objectives.¹

Earlier observational comparisons suggested that adding oxygen-directed treatment might improve survival, although historical controls and differences in care limited causal interpretation. BOOST-II subsequently demonstrated that a protocol informed by PbtO_2 could reduce measured brain hypoxia. Its clinical findings were encouraging but did not establish improved long-term recovery. The next question was therefore whether adding the monitor and acting on its measurements produced a worthwhile patient benefit.^{2,3}

That question also requires consideration of harm. Inserting another intracerebral catheter introduces procedural risks and technical failures, while responding to low readings may expose patients to additional treatment. OXY-TC tested an early, five-day combined ICP- and PbtO_2 -guided strategy against active ICP-guided care, using

six-month functional outcome rather than correction of the monitored variable as the primary endpoint.¹

Synopsis

OXY-TC was an investigator-initiated, multicentre, parallel-group, open-label superiority trial conducted at 25 French tertiary referral centres with continuous neurosurgical availability. Recruitment ran from 15 June 2016 to 17 April 2021. Funding came from the French National Program for Clinical Research and La Fondation des Gueules Cassées; Integra Lifesciences supplied monitoring equipment and probes. Funders had no role in study design, data analysis, interpretation or writing.^{1,4}

Patients aged 18–75 years were eligible after severe non-penetrating TBI, with a best prehospital Glasgow Coma Scale (GCS) score of 3–8, a motor score of 1–5 and an indication for ICP monitoring. Sedation and mechanical ventilation were expected to exceed 48 hours. Patients had to be stabilised, with a $\text{PaO}_2/\text{FiO}_2$ ratio above 150 and mean arterial pressure above 70 mmHg; intracerebral monitoring had to begin within 16 hours of injury. Important exclusions included GCS 3 with bilateral fixed dilated pupils, prior decompressive craniectomy, quadriplegia, coagulation disorders preventing probe insertion, temperature below 34°C, anticipated survival below 24 hours, post-traumatic cardiac arrest, and conditions preventing reliable follow-up. Consent came from a relative or legal surrogate, with emergency enrolment and subsequent consent where permitted.¹

A secure website allocated patients 1:1 using variable-sized blocks, stratified by centre and age below or at least 50 years. Allocation was concealed before randomisation. Treating clinicians could see the monitoring strategy, but central outcome assessors and statisticians were masked. Clinicians received training in interpreting and responding to PbtO_2 measurements.¹

Both groups received an ICP probe and protocolised care for five days, targeting ICP at or below 20 mmHg. Basic management included sedation, analgesia and ventilation, PaCO_2 35–40 mmHg, PaO_2 80–140 mmHg, CPP 60–70 mmHg, glucose 6–10 mmol/L, sodium 140–150 mmol/L, haemoglobin 7–10 g/dL, temperature 36–38°C and 15° head elevation. If ICP exceeded 20 mmHg, second-tier options included deeper sedation, vasopressors to increase CPP above 70 mmHg, PaCO_2 30–35 mmHg, osmotherapy, external ventricular drainage, neuromuscular blockade and temperature control at 35–37°C. Refractory intracranial hypertension prompted third-tier treatment with hypothermia at 33–35°C, barbiturates or secondary decompressive craniectomy.^{1,4}

The dual-monitoring group additionally received a Licox oxygen and temperature catheter, inserted 20–25 mm below the dura into tissue unaffected by the injury, or the right frontal lobe in diffuse injury. An oxygen challenge within two hours checked probe function, seeking a 100–300% PbtO₂ increase with FiO₂ 1.0. When PbtO₂ fell below 20 mmHg, the algorithm specified interventions in sequence: PaO₂ 100–150 mmHg, PaCO₂ 35–45 mmHg, temperature 35–37°C, CPP 60–100 mmHg, cardiac index above 2.5 L/min/m², haemoglobin 9–12 g/dL, and finally PaO₂ above 150 mmHg. With both intracranial hypertension and low PbtO₂, ICP-directed treatment preceded oxygen-directed escalation. These were protocol targets, not separately randomised comparisons of oxygen, transfusion or blood-pressure thresholds.^{1,4}

The primary outcome was an unfavourable Extended Glasgow Outcome Scale (GOSE) score of 1–4 at six months, assessed through a central structured telephone interview. This category included death, vegetative state and lower or upper severe disability; scores 5–8 represented moderate disability through good recovery. The original MRI-based primary endpoint, brain injury volume at days 6–10, was replaced with GOSE after technical difficulties, with approval on 7 February 2018. Secondary outcomes included 12-month GOSE, Disability Rating Scale (DRS) and Functional Independence Measure (FIM) scores at six and 12 months, day-28 survival, therapeutic intensity and critical intracranial events.^{1,4}

The planned sample of 300, allowing for 14 patients lost to follow-up, provided 80% power at a two-sided 5% significance level to detect a 30% relative reduction from an expected 55% unfavourable outcome rate: approximately 17 percentage points. An amendment authorised 20 additional participants to offset losses to the analysable population. No interim efficacy analysis was performed. Logistic regression adjusted for age and centre; odds ratios below 1 favoured dual monitoring. The modified intention-to-treat population excluded consent withdrawals and specified eligibility violations. Missing primary outcomes were imputed using chained equations for a sensitivity analysis, alongside a per-protocol analysis. Secondary testing was not adjusted for multiplicity.^{1,4}

Of 2578 patients screened, 318 were randomised: 160 to ICP-only and 158 to dual monitoring. Eight consent withdrawals and 19 protocol violations left 144 and 147 patients, respectively, in the modified intention-to-treat population. Six-month GOSE was available for 132 and 139, leaving 20 missing outcomes. The per-protocol population comprised 139 ICP-only and 125 dual-monitoring patients. Protocol deviations affected five versus 22 patients, including catheter dysfunction, absent or insufficient monitoring and late initiation.^{1,4}

Median ages were 38 and 40 years; 74% were male in each group. Road traffic accidents accounted for approximately two-thirds of injuries, and around half had major extracranial injuries. Median predicted probabilities of poor outcome were 0.57 in both groups. Admission ICP was 11 versus 12 mmHg and $\text{PaO}_2/\text{FiO}_2$ was 400 versus 390; median PbtO_2 in the dual-monitoring group was 14 mmHg. Noradrenaline was used in 63% versus 73%, respectively, although the overall baseline profiles were broadly similar.¹

All patients received second-tier treatment. Third-tier treatment was used in 49/142 ICP-only patients (35%) and 48/141 dual-monitoring patients (34%). In the latter group, 101/137 patients (74%) received at least one intervention to correct brain hypoxia, and 38/132 (29%) experienced PbtO_2 below 10 mmHg for at least 30 minutes. Median PbtO_2 rose during early monitoring, but oxygen measurements were unavailable in controls. Median maximum therapeutic intensity scores were 10 versus 11, with no significant difference, and daily scores declined similarly. Group-level pressure measurements were generally within the intended ranges.^{1,4}

In the complete-case primary analysis, unfavourable six-month outcome occurred in 70/132 ICP-only patients (53%) and 72/139 dual-monitoring patients (52%): adjusted odds ratio 0.9 (95% CI, 0.6–1.6; $P = 0.83$). After imputation, estimated rates were 51% versus 52%: odds ratio 1.0 (95% CI, 0.6–1.7; $P = 0.95$). The abstract presents this imputed estimate. Per-protocol results were also inconclusive: 66/127 (52%) versus 58/119 (49%); odds ratio 0.9 (95% CI, 0.5–1.6; $P = 0.62$).^{1,4}

At 12 months, unfavourable outcome occurred in 61/126 patients (48%) versus 58/135 (43%): odds ratio 0.8 (95% CI, 0.5–1.3; $P = 0.34$). DRS and FIM scores did not differ significantly. Day-28 mortality was 28/144 (19%) versus 25/145 (17%): odds ratio 0.8 (95% CI, 0.5–1.5; $P = 0.53$). Twelve-month survival status was available for 275 patients; there were 33 versus 34 deaths, with a reported adjusted hazard ratio of 1.09 (95% CI, 0.67–1.77; $P = 0.72$). Median intensive care unit (ICU) stays were 20 versus 22 days ($P = 0.18$).¹

A post hoc subgroup of 55 patients with admission ICP at least 20 mmHg had unfavourable outcomes in 25/28 controls (89%) versus 14/27 dual-monitoring patients (52%): odds ratio 0.13 (95% CI, 0.02–0.86; $P = 0.034$). The interaction P value comparing treatment effects by admission ICP category was 0.054. Other exploratory severity subgroups did not show a difference.^{1,4}

Catheter dysfunction occurred in 2/144 controls (1%) versus 12/147 dual-monitoring patients (8%; $P = 0.011$), and catheter-related intracerebral haematoma in none versus

six (4%; $P = 0.030$). Serious trial-related adverse events affected two versus seven patients: odds ratio 3.5 (95% CI, 0.8–15.5; $P = 0.090$). No statistically significant differences were identified in ICU-related complications, although ventilator-associated pneumonia occurred in 31% versus 42% and pulmonary embolism in 4% versus 8%.¹

Critique

OXY-TC did not demonstrate that routinely adding its oxygen-monitoring strategy improved six-month neurological outcome. However, the intervention was a monitor coupled to a particular treatment algorithm, not a test of whether cerebral oxygenation matters. Patients received additional information and a sequence of responses to it. The result therefore concerns the net clinical effect of that package, including its procedural risks, treatment consequences and practical delivery, rather than every possible use of PbtO_2 monitoring.^{1,4}

The comparator was active and substantial. Both groups received structured ICP and CPP management, all received second-tier treatment, and approximately one-third required third-tier therapy. Dual monitoring had to add benefit beyond this care. Conversely, a largely normal admission ICP does not mean the population lacked an oxygenation problem: median initial PbtO_2 was 14 mmHg, and most monitored patients subsequently received oxygen-directed interventions. Describing the trial simply as an investigation of patients who were insufficiently ill would therefore miss its actual physiology.¹

The intervention also bundled several distinct therapeutic choices. Adjusting inspired oxygen, ventilation, perfusion pressure, cardiac output and haemoglobin need not have identical physiological consequences. OXY-TC cannot identify the most useful component or determine whether another sequence would perform better. Similar aggregate therapeutic intensity scores do not establish identical oxygen exposure, vasopressor doses or transfusion practice. The supplied results do not provide sufficiently detailed comparative exposure data to determine which interventions produced the greatest separation or whether their benefits and harms offset one another.^{1,4}

A more fundamental limitation is that PbtO_2 was not recorded in the control group. Rising values in the dual-monitoring arm show its observed course, not the reduction in brain hypoxia attributable to randomisation. Unlike BOOST-II, OXY-TC cannot establish a between-group difference in the putative mediator. Consequently, the neutral clinical result cannot distinguish failure to reduce hypoxia from failure of hypoxia correction to improve recovery. Inserting blinded probes in controls would have answered that

physiological question, but would also have exposed controls to the procedure whose additional risk is part of the clinical comparison.^{1,3}

The focal nature of the measurement adds another uncertainty. The catheter sampled relatively uninjured tissue rather than the entire injured brain. A normal value at that site cannot demonstrate that more severely affected regions received sufficient oxygen, while a treatment-related increase does not itself establish restored metabolism. These are limitations of interpreting the signal, not proof that probe location explains the result. Similarly, hourly recordings and six-hourly summaries cannot fully describe the duration and depth of intracranial insults. Acceptable group medians can coexist with clinically important episodes, as illustrated by severe hypoxia in 29% of monitored patients.^{1,4}

Methodologically, central concealed randomisation, stratification, multicentre recruitment and masked structured outcome assessment support the comparison. Blinding clinicians was impractical, but independent assessment reduced the risk that knowledge of allocation influenced GOS classification. It could not eliminate differences in bedside decisions or subsequent care. The primary outcome appropriately incorporated survival and disability, although its dichotomy grouped markedly different states together. An improvement within a category would not alter the primary endpoint; the published category distributions are therefore informative even without a positive binary comparison.^{1,4}

The change in primary endpoint merits a proportionate interpretation. The protocol explains that, after 28 patients had been randomised, more than 60% of the intended MRI endpoints were missing because of safety or logistical constraints. Moving to a patient-centred functional outcome was defensible and transparently documented. This is not, by itself, evidence of choosing an outcome because its results were favourable. Nevertheless, the detailed supplied statistical analysis plan is dated March 2023, after recruitment had ended. Its availability improves auditability but does not establish that every analysis was specified before enrolment; the severity subgroup analyses are explicitly described as outside the original protocol.⁴

The analysis also stopped short of including all randomised participants. Twenty-seven were removed before the modified intention-to-treat analysis, and a further 20 lacked the primary measurement. Imputation addressed the latter group, not the former. Similar complete-case and imputed estimates are reassuring under the imputation assumptions, but cannot prove missing outcomes were inconsequential. The per-protocol analysis removed substantially more dual-monitoring patients. Excluding technical failure may explore efficacy when monitoring is delivered, but also conditions

on post-randomisation events and removes a disadvantage relevant to real-world implementation.^{1,4}

Precision remains important despite the almost identical event rates. The planned 17-percentage-point improvement was ambitious, and the confidence interval still includes potentially important benefit and harm. The trial failed to demonstrate superiority; it did not establish equivalence. Fewer complete outcomes than intended increased uncertainty, but a generic claim of insufficient power should not obscure the actual finding: neither the primary analysis nor its sensitivity analyses showed an overall advantage. The numerically lower unfavourable outcome rate at 12 months also remained uncertain and was a secondary comparison.¹

The high-ICP subgroup is the most tempting route to a positive interpretation. Its large apparent benefit concerns only 55 patients, identified through post hoc analyses of several severity characteristics. The interaction P value of 0.054 does not cross the trial's conventional 0.05 threshold. More importantly, a significant comparison within one subgroup is not sufficient to establish that treatment works differently across subgroups. Randomisation is not automatically lost by examining a baseline subgroup, but small numbers, analytical selection and multiple testing weaken a treatment-selection claim. These data justify prospective testing of an enriched population, not a validated indication for routine dual monitoring whenever admission ICP is at least 20 mmHg.^{1,4}

The catheter-related harm signal must remain part of that judgement. Six intracerebral haematomas and more frequent technical failures are clinically relevant when a neurological benefit has not been demonstrated. Their contribution to long-term disability cannot be determined from the supplied data. Nor do nonsignificant comparisons of pneumonia or pulmonary embolism establish equal systemic safety. The appropriate conclusion is an uncertain benefit–harm balance, not that every complication arose from oxygen-directed treatment or that failure to reach statistical significance makes the strategy harmless.¹

Finally, applicability is limited to patients sufficiently stable for early monitoring, aged no more than 75 years, without prior decompressive craniectomy or the other major exclusions. Only about one in eight screened patients was randomised. The five-day treatment window captured early management but not the full ICU admission, and does not settle whether longer monitoring, a different algorithm or another selected population would benefit. These remain distinct research questions. They do not provide grounds to assume benefit where the tested strategy failed to demonstrate it.^{1,4}

Body of Evidence

Early comparative evidence was encouraging but vulnerable to confounding. Stiefel and colleagues compared 28 patients receiving brain oxygen-directed care with 25 historical controls matched for several baseline characteristics. Both groups targeted ICP below 20 mmHg and CPP above 60 mmHg; the additional oxygen target was above 25 mmHg. Mortality was 44% in controls and 25% with oxygen-directed treatment. This small observational comparison supported the hypothesis, but matching could not separate the additional monitoring from changes in care over time or other unmeasured differences.²

Lin and colleagues subsequently conducted a small randomised study across six hospitals in Taiwan. Fifty patients with GCS below 12 received ICP-guided treatment alone (27 patients) or PbtO₂-guided management (23 patients), targeting brain tissue oxygen pressure above 20 mmHg. Patients with severe systemic injury and several other major conditions were excluded. Better survival was reported at three and six months, but the small population and inclusion of moderate as well as severe TBI limit direct comparison with OXY-TC. This was supportive preliminary evidence, not a precise estimate of the effect of routine monitoring in the broader severe-TBI population.⁵

BOOST-II randomised 119 patients at ten US centres, 62 to ICP-guided and 57 to combined ICP- and PbtO₂-guided treatment. Crucially, both groups received oxygen probes, with the control readings concealed from clinicians. Continuous recordings demonstrated reductions of 66% in hypoxia duration and 72% in its average depth, without a comparable improvement in ICP control. Six-month GOSE was available for 106 patients; mortality was 34% with ICP-only treatment and 25% with dual guidance, but clinical differences were not statistically established. The trial was designed to demonstrate physiological efficacy and feasibility rather than definitive neurological benefit.³

The 2020 Seattle International Severe Traumatic Brain Injury Consensus Conference (SIBICC) addressed how to manage adults who already had both ICP and brain oxygen monitors. Forty-two experts developed tiered algorithms for raised ICP, low brain oxygen, or both. This was consensus guidance, not randomised evidence that inserting the additional monitor improves outcome. Its explicit distinction between a practical management framework and a proven standard of care is important: the availability of a coherent response to a low reading does not resolve whether routine deployment benefits patients.⁶

OXY-TC added a larger randomised comparison with functional outcome as its primary endpoint and a control group not exposed to the additional oxygen catheter. Its neutral

result therefore carries more direct information about routine clinical adoption than a physiological improvement alone. It does not simply overturn BOOST-II: the trials differed in the question prioritised, control instrumentation and treatment delivery. One demonstrated that the monitored hypoxia burden could be reduced; the other did not establish that its monitoring-and-treatment package improved six-month recovery. Those findings can coexist without either study being dismissed.^{1,3}

At the end of 2023, the evidence did not establish a neurological benefit sufficient to justify routine PbtO₂ monitoring for every eligible patient with severe TBI. Small positive studies, a convincing physiological signal and expert agreement about how to respond to abnormal values supported further investigation, but were not substitutes for clinical efficacy. OXY-TC also made technical reliability and procedural harm harder to treat as peripheral concerns. Its investigators identified BOOST-3 and BONANZA as ongoing trials intended to strengthen the outcome evidence. Whether a reproducibly defined, higher-risk subgroup benefits remained an important unanswered question.^{1,6}

Should we routinely add PbtO₂ monitoring in severe TBI?

Not routinely. OXY-TC did not demonstrate better neurological outcomes and reported more catheter-related complications.

References

1. Payen JF, Launey Y, Chabanne R, Gay S, Francony G, Gergele L, et al; OXY-TC trial collaborators. Intracranial pressure monitoring with and without brain tissue oxygen pressure monitoring for severe traumatic brain injury in France (OXY-TC): an open-label, randomised controlled superiority trial. *Lancet Neurol* 2023;22(11):1005–1014.
2. Stiefel MF, Spiotta A, Gracias VH, Garuffe AM, Guillaumondegui O, Maloney-Wilensky E, et al. Reduced mortality rate in patients with severe traumatic brain injury treated with brain tissue oxygen monitoring. *J Neurosurg* 2005;103(5):805–811.
3. Okonkwo DO, Shutter LA, Moore C, Temkin NR, Puccio AM, Madden CJ, et al. Brain Oxygen Optimization in Severe Traumatic Brain Injury Phase-II: A Phase II Randomized Trial. *Crit Care Med* 2017;45(11):1907–1914.

- 
4. Payen JF, Launey Y, Chabanne R, Gay S, Francony G, Gergele L, et al; OXY-TC trial collaborators. Supplementary appendix: therapeutic algorithms, additional results, trial protocol and statistical analysis plan. Supplement to: Lancet Neurol 2023;22(11):1005–1014.
 5. Lin CM, Lin MC, Huang SJ, Chang CK, Chao DP, Lui TN, et al. A Prospective Randomized Study of Brain Tissue Oxygen Pressure-Guided Management in Moderate and Severe Traumatic Brain Injury Patients. Biomed Res Int 2015;2015:529580.
 6. Chesnut R, Aguilera S, Buki A, Bulger E, Citerio G, Cooper DJ, et al. A management algorithm for adult patients with both brain oxygen and intracranial pressure monitoring: the Seattle International Severe Traumatic Brain Injury Consensus Conference (SIBICC). Intensive Care Med 2020;46(5):919–929.

TENSION

Bendszus M, Fiehler J, Subtil F, Bonekamp S, Aamodt AH, Fuentes B, et al. Endovascular thrombectomy for acute ischaemic stroke with established large infarct: multicentre, open-label, randomised trial. Lancet 2023;402(10414):1753–1763.

Introduction

A large established infarct presents a difficult boundary for reperfusion treatment. Reopening an occluded artery may preserve remaining viable tissue, but cannot be assumed to reverse established injury. When extensive ischaemic changes are already visible, the prospects for independence are poor and the risks of haemorrhage and swelling become particularly important. The pivotal thrombectomy trials predominantly enrolled patients with limited or moderate infarction, leaving uncertainty about patients whose imaging suggested that much of the threatened territory was already damaged.^{1,2}

The Alberta Stroke Program Early Computed Tomographic Score (ASPECTS) describes the extent of ischaemic changes in the middle cerebral artery territory. Scores range from 10 to 0, with lower scores indicating more extensive involvement. The distinction between a prognostic marker and a treatment-selection rule is important: a patient likely to have a poor outcome without treatment may nevertheless benefit from treatment. Conversely, a low ASPECTS does not quantify the remaining salvageable tissue or establish that reperfusion will be worthwhile.¹

Extending the treatment window had previously depended on selecting patients with favourable clinical or perfusion-imaging profiles. DEFUSE 3, for example, required a limited infarct core and a substantial mismatch between the core and hypoperfused tissue.³ More recent randomised trials challenged exclusion solely because the infarct was large. However, MRI or perfusion imaging featured prominently in their selection pathways, whereas many hospitals assess infarct extent visually on non-contrast CT.^{1,4,5}

TENSION therefore addressed both a clinical uncertainty and a practical barrier: could thrombectomy improve outcomes in patients with ASPECTS 3–5, using locally interpreted standard imaging rather than a requirement for perfusion-based selection? Its treatment window extended to 12 hours, but its central question concerned established infarction, not simply late presentation.¹

Synopsis

TENSION was an investigator-initiated, international, open-label, parallel-group randomised trial with blinded endpoint assessment. It compared endovascular thrombectomy plus medical treatment with medical treatment alone in anterior-circulation large-vessel occlusion with an established large infarct. The trial was conducted across 40 European hospitals and one Canadian site, with recruitment from 17 July 2018 to 21 February 2023. Funding came from the European Union Horizon 2020 programme; the funder had no role in design, conduct, analysis or reporting.¹

Eligible patients were adults with an intracranial distal internal carotid artery or M1 middle cerebral artery occlusion confirmed by CT or MR angiography, and a locally assessed ASPECTS of 3–5 on non-contrast CT or diffusion-weighted MRI. Pre-stroke modified Rankin Scale (mRS) had to be 0–2 and National Institutes of Health Stroke Scale (NIHSS) below 26. Randomisation was required within 11 hours of symptom onset or last known well, with thrombectomy expected to finish within 12 hours. Important exclusions included intracranial haemorrhage, mass effect, vascular anatomy preventing thrombectomy, and high-grade extracranial stenosis expected to require acute stenting.^{1,2}

Allocation was central and web-based, using variable blocks of two or four, stratified by presentation before six hours versus 6–11 hours and NIHSS of 18 or less versus 19–25. Consent arrangements followed national requirements, including representative or independent-physician consent where applicable. Treating teams were unblinded. Follow-up assessments were performed by trained personnel unaware of allocation, while a blinded central laboratory reviewed imaging. Investigators completed ASPECTS training before recruitment, but enrolment depended on the local assessment.^{1,2}

Thrombectomy technique was discretionary: aspiration, stent retrieval or both, with or without balloon protection. Both groups received guideline-based medical management, including intravenous thrombolysis when indicated and supportive stroke care. Thus, the comparison tested adding thrombectomy, not replacing medical treatment or withholding thrombolysis. Assessments occurred at baseline, approximately 24 hours, seven days or discharge, and 90 days. Telephone mRS assessment was permitted when an in-person visit was impractical.^{1,2}

The primary endpoint was the ordinal distribution of mRS at 90 days, with a 14-day assessment window. This scale ranges from no symptoms to death; scores 0 and 1 were combined for analysis because of sparse observations. Secondary outcomes included independence, defined as mRS 0–2, and mRS 0–3, which includes patients requiring some help but able to walk unassisted. Decompressive surgery, infarct volume at 24 hours,

quality of life, anxiety and depression were also assessed. Safety outcomes included mortality, serious adverse events and symptomatic intracranial haemorrhage. The latter required an increase of at least four NIHSS points, or two points in a subcategory, together with parenchymal haemorrhage type 2: haemorrhage involving more than 30% of the infarct with substantial mass effect, or bleeding remote from the original infarct.¹

The sample-size calculation targeted a common odds ratio of 1.5 for improved mRS, with 80% power and a one-sided significance level of 0.025. The statistical plan increased an initial requirement of 620 patients to 633 for the sequential design and then to 665 allowing for attrition. Two interim analyses, after approximately one-third and two-thirds completed follow-up, permitted stopping for efficacy or futility. The first efficacy boundary was a test statistic of 3.11. Efficacy analyses retained randomised allocation; safety analyses classified patients by whether an endovascular procedure had been initiated.^{1,2}

The primary model adjusted for the randomisation strata. Seven missing mRS outcomes were multiply imputed using age, baseline and discharge NIHSS, and treatment allocation. Supplementary Table S4 specifies separate coefficients across outcome thresholds for the time stratum because its proportional-odds assumption was not met, while retaining a common treatment odds ratio. Per-protocol and best-case/worst-case analyses assessed robustness. Secondary outcomes and subgroup comparisons were not adjusted for multiplicity.^{1,2}

Recruitment stopped after the first planned interim analysis of 222 patients crossed the efficacy boundary. Its sequential treatment estimate was 3.05 (one-sided 97.5% lower confidence bound 1.82; $P < 0.0001$). By then, 253 had been randomised, 125 to thrombectomy and 128 to medical treatment. All remained in the primary analysis. Three control patients underwent thrombectomy, giving safety populations of 128 and 125. The per-protocol populations were 97 and 89 after excluding important clinical or imaging deviations. Primary outcome observations were available for 124 and 122 patients before imputation, using the denominators in Supplementary Table S4.^{1,2}

Median age was 73 versus 74 years and median NIHSS 19 versus 18. Women represented 45% versus 52%. Approximately 57% had a known onset and more than half had been transferred to a thrombectomy centre. CT was the qualifying modality in 208 patients, or 82%, and MRI in 45. Central review subsequently classified 50 of 252 assessable baseline scans outside ASPECTS 3–5. Intravenous alteplase was given to 49/125 patients versus 44/128, or 39% versus 34%.^{1,2}

Actual thrombectomy was performed in 123 intervention patients; spontaneous recanalisation or distal embolus migration prevented retrieval in the other two. Successful reperfusion, defined as modified Thrombolysis in Cerebral Ischaemia grade 2b or better, occurred in 104/125 patients, or 83%. Combined aspiration and stent retrieval predominated, with a median of three attempts. General anaesthesia was used in 44%. The main treatment table gives a median onset-to-groin-puncture interval of 4.2 hours and a randomisation-to-recanalisation interval of 2.4 hours.^{1,2}

At 90 days, median mRS was 4 (IQR 3–6) versus 6 (4–6). The adjusted common odds ratio favouring thrombectomy was 2.58 (95% CI 1.60–4.15; $P=0.0001$). Independence occurred in 21/124 versus 3/122 patients, or 17% versus 2% (adjusted odds ratio 7.16; 95% CI 2.12–24.21). An mRS of 0–3 occurred in 39/124 versus 16/122, or 31% versus 13% (adjusted odds ratio 2.84; 95% CI 1.48–5.47). These secondary comparisons were nominally significant, without multiplicity correction.^{1,2}

Mean infarct volume at 24 hours was 205.8 versus 227.7 mL (adjusted ratio 0.93; 95% CI 0.80–1.08; $P=0.35$). Decompressive surgery occurred in 11/125 versus 9/128 patients. Available quality-of-life responses favoured thrombectomy, including median EQ-5D indices of 0.6 versus 0.4. In the separately defined safety analysis, mortality censored at 90 days was 49/122 versus 63/123, or 40% versus 51% (adjusted hazard ratio 0.67; 95% CI 0.46–0.98; $P=0.038$). Symptomatic intracranial haemorrhage occurred in 7/128 versus 6/125 patients, with three versus four fatal events. At least one serious adverse event occurred in 71/128 versus 88/125 patients. Procedure-related serious adverse events affected approximately 7% of thrombectomy recipients, without a reported procedure-related death.^{1,2}

Critique

TENSION demonstrated a clinically important improvement in the distribution of disability, not a high likelihood of recovery to independence. The distinction matters when communicating its results. A common odds ratio of 2.58 does not mean that independence became 2.58 times as likely, nor that each treated patient improved by two mRS categories. The observed independence rate was only 17% after thrombectomy, and the median treated patient had moderately severe disability. Nevertheless, increasing the proportion able to walk without assistance from 13% to 31% represents a meaningful benefit in this severely affected population.¹

The outcome distribution also argues against describing the result simply as survival with extreme disability. In the randomised groups, the proportions with mRS 5 and with death were both lower after thrombectomy, while more patients reached less dependent



states. These are population-level comparisons, not evidence identifying which individuals were saved or what their alternative outcomes would have been. The supplementary analysis restricted to survivors cannot establish an independent causal effect on disability among people who would survive either treatment: survival may itself be affected by treatment, so conditioning on it can destroy the comparability created by randomisation.²

The imaging strategy is TENSION's most distinctive contribution. Its benefit was demonstrated in a population selected predominantly by ordinary CT and angiography, without requiring a perfusion mismatch or automated core-volume threshold. This removes an important evidential objection to treating an otherwise eligible patient where advanced imaging is unavailable. It does not demonstrate that CT selection is equivalent or superior to MRI or perfusion imaging, because imaging strategies were not randomised. Nor does it remove the need for vascular imaging, reliable interpretation or an experienced thrombectomy service.^{1,2}

ASPECTS variability was substantial: central review placed about one-fifth of patients outside the intended range, including 38 with scores of 0–2. Retaining them in the primary analysis was appropriate because treatment had been allocated using the local assessment. Excluding them retrospectively would change the clinical question from whether locally selected patients benefit to whether centrally confirmed patients benefit. The similar per-protocol estimate is reassuring, but this smaller selected population does not replace the intention-to-treat result. Importantly, the very imprecise estimate in the centrally classified ASPECTS 0–2 subgroup, odds ratio 1.51 (95% CI 0.44–5.19), does not independently validate routine treatment of that category.^{1,2}

More control patients had centrally assessed ASPECTS 0–2, raising the possibility that a chance prognostic imbalance contributed to the difference. A post hoc analysis adjusting for baseline ASPECTS still favoured thrombectomy, with an odds ratio of 3.12 (95% CI 1.61–6.06). This weakens that explanation without making the exploratory model definitive. Central allocation and blinded outcome and imaging assessments provide stronger protection against systematic bias than an assertion that all baseline variables were perfectly balanced.^{1,2}

Early stopping requires a more nuanced interpretation than either dismissing the result or overlooking the reduced information. Sequential efficacy monitoring was specified before recruitment, including a stringent first stopping boundary; the primary finding was therefore not generated by an unplanned inspection followed by stopping at $P < 0.05$. However, stopping after 253 rather than 665 patients leaves the magnitude of

benefit, uncommon harms and subgroup effects less securely estimated. The interim treatment estimate was larger than that in the full cohort, consistent with the need not to regard the first estimate as a precise forecast for practice.^{1,2}

There is also an important reporting distinction. The final statistical analysis plan, dated July 2023, after recruitment ended, describes cessation in response to external evidence. It specifies a sequential test for the 222-patient interim dataset and, if that boundary was crossed, estimation without an additional test for the entire recruited cohort. The main publication describes stopping for efficacy and prominently reports $P=0.0001$ for all 253 patients. Supplementary Table S3 provides a positive sequential interim analysis. The evidence for benefit remains strong, but the chronology and the difference between the planned and reported full-cohort testing warrant clarification; the two analyses should not be presented as independent confirmations.^{1,2}

The analysis plans also differ in their handling of the mRS categories. Version 1 pooled scores 5 and 6, whereas the final plan pooled scores 0 and 1. An undated version-2 amendment records this as a correction. Separating death from severe disability is clinically defensible, but changes what the ordinal analysis distinguishes. The timing and rationale merit confirmation rather than an assumption that the change was either innocuous or outcome-driven.²

Missing primary outcomes are unlikely to explain the principal finding. Although missingness was asymmetric, one versus six patients, the deliberately unfavourable sensitivity analysis still favoured thrombectomy (odds ratio 2.21; 95% CI 1.39–3.53). The detailed supplementary model also addresses non-proportionality in the time covariate rather than assuming every model term meets the same restriction. These strengths do not eliminate the consequences of open-label care: decisions about escalation, limitation of treatment and rehabilitation could still differ. The supplied report does not provide enough detail to quantify such differences, and their existence should not be assumed.^{1,2}

The mortality finding is supportive but less definitive than the primary ordinal result. It was one of several unadjusted comparisons, its confidence interval narrowly excluded no effect, and the published hazard ratio comes from an as-treated safety population rather than the randomised efficacy comparison. Reassigning three crossover patients is particularly relevant when interpreting survival. The mRS distribution itself incorporates deaths in the allocated groups, but its death counts and assessment window should not be interchanged with the separately reported survival analysis. Consequently, the strongest conclusion is improved overall functional outcome, with a possible survival advantage requiring proportionate qualification.^{1,2}



Similarly, comparable haemorrhage counts do not establish the absence of procedural harm. Only 13 symptomatic events occurred, the definition required substantial parenchymal haemorrhage plus neurological deterioration, and approximately 7% of intervention patients had procedure-related serious adverse events. The balance nevertheless favoured thrombectomy for the primary patient-centred outcome. Lack of a clear reduction in infarct volume does not negate that result: the imaging comparison was imprecise and was not a direct measurement of rescued tissue. TENSION neither demonstrates reversal of irreversibly infarcted brain nor identifies the mechanism responsible for the functional benefit.^{1,2}

Finally, favourable quality-of-life comparisons were based on available responses, not complete measurements from all randomised patients. Differential survival and ability to respond limit their interpretation. Applicability also remains bounded by the trial's eligibility: NIHSS below 26, pre-stroke mRS no greater than 2, absence of important mass effect, proximal anterior-circulation occlusion and treatment completed within 12 hours. The small older-age and very-low-ASPECTS subgroups cannot reliably establish separate treatment rules. No overall screening denominator was provided, so the proportion of potentially eligible patients actually recruited is uncertain. These limitations refine where the result applies; they do not erase the benefit demonstrated in the enrolled population.^{1,2}

Body of Evidence

The earlier thrombectomy evidence established a substantial treatment effect in proximal anterior-circulation occlusion but offered much less direct evidence for extensive established infarction. The original TENSION protocol highlights the predominance of favourable prognostic features in the early trials, including limited ischaemic injury. MR CLEAN did not require a minimum ASPECTS, but allowing some patients with extensive infarction is not equivalent to a trial specifically designed and powered for that population. TENSION was designed to address that gap rather than repeat the original broad demonstration that thrombectomy can work.²

DEFUSE 3 extended the time window in a selected population. Across 38 US centres, 182 patients were randomised to thrombectomy plus medical treatment or medical treatment alone between six and 16 hours after last known well. Selection required a core below 70 mL, a hypoperfused-to-core ratio of at least 1.8 and at least 15 mL of potentially salvageable tissue. The trial stopped early; thrombectomy improved the 90-day mRS distribution (odds ratio 2.77; $P < 0.001$), and independence occurred in 45% versus 17%. Its central lesson was that favourable tissue characteristics could identify

late responders. TENSION addressed a different selection problem: visible extensive infarction, without a mandatory perfusion mismatch.³

RESCUE-Japan LIMIT provided direct randomised evidence for large infarcts. It enrolled 203 patients with ASPECTS 3–5, assigning 101 to thrombectomy and 102 to medical care. Treatment was within six hours of last known well, or within 24 hours when specified MRI findings supported eligibility. The primary outcome, mRS 0–3 at 90 days, occurred in 31.0% versus 12.7% (relative risk 2.43; 95% CI 1.35–4.37; $P=0.002$). Any intracranial haemorrhage was more frequent after thrombectomy, 58.0% versus 31.4%; this should not be mistaken for the symptomatic-haemorrhage rate. Approximately 86% were selected using MRI, an important difference from TENSION's predominantly CT-based population.^{1,4}

SELECT2 subsequently randomised 352 patients, 178 to thrombectomy and 174 to medical care, within 24 hours. Large core was defined by ASPECTS 3–5 or a core volume of at least 50 mL on CT perfusion or diffusion-weighted MRI. Recruitment stopped early for efficacy. The 90-day mRS distribution favoured thrombectomy (generalised odds ratio 1.51; 95% CI 1.20–1.89; $P<0.001$); independence occurred in 20% versus 7% (relative risk 2.97; 95% CI 1.60–5.51). Mortality was similar. Symptomatic intracranial haemorrhage occurred in one versus two patients, although procedural vascular complications occurred in the intervention group. SELECT2 therefore supported functional benefit across a different large-core definition and a longer window, without establishing a survival benefit.⁵

ANGEL-ASPECT added a 456-patient comparison from 46 Chinese centres, treating within 24 hours. Eligibility used ASPECTS 3–5 or a specified core-volume range of 70–100 mL. The 90-day mRS distribution favoured endovascular therapy, with a generalised odds ratio of 1.37. Any intracranial haemorrhage was more frequent, without a clear difference in symptomatic haemorrhage. Its contribution was corroboration in another healthcare setting, but its eligibility criteria and permitted endovascular treatments should not be collapsed into TENSION's simpler visual-score rule. Alongside RESCUE-Japan LIMIT and SELECT2, it provided a third positive randomised comparison before TENSION was published.^{1,6,7}

Taken together, these trials shift the question from whether a large infarct necessarily precludes benefit to which patients within that broad category should be treated. Their findings are mutually supportive, but differing populations, imaging definitions, treatment windows and ordinal effect measures prevent a simple ranking of benefit. In particular, SELECT2 allowed an ASPECTS-based eligibility route; it would be inaccurate to imply that every preceding patient required perfusion selection. TENSION's additional

contribution is positive randomised evidence using predominantly non-contrast CT, alongside angiography, in patients treatable within 12 hours. Even with treatment, severe disability and death remained common, making the likely range of outcomes central to discussions with patients and families.^{1,4,5}

Should we offer thrombectomy for large infarcts with ASPECTS 3–5?

Yes, in otherwise eligible patients treatable within 12 hours. TENSION improved functional outcomes, although most patients remained dependent or died.

References

1. Bendszus M, Fiehler J, Subtil F, Bonekamp S, Aamodt AH, Fuentes B, et al. Endovascular thrombectomy for acute ischaemic stroke with established large infarct: multicentre, open-label, randomised trial. *Lancet* 2023;402(10414):1753–1763.
2. Bendszus M, Fiehler J, Subtil F, et al. Supplementary appendix to: Endovascular thrombectomy for acute ischaemic stroke with established large infarct: multicentre, open-label, randomised trial. *Lancet* 2023;402:1753–1763. Includes trial protocols and statistical analysis plans.
3. Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, et al. Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging. *N Engl J Med* 2018;378(8):708–718.
4. Yoshimura S, Sakai N, Yamagami H, Uchida K, Beppu M, Toyoda K, et al. Endovascular Therapy for Acute Stroke with a Large Ischemic Region. *N Engl J Med* 2022;386(14):1303–1313.
5. Sarraj A, Hassan AE, Abraham MG, Ortega-Gutierrez S, Kasner SE, Hussain MS, et al. Trial of Endovascular Thrombectomy for Large Ischemic Strokes. *N Engl J Med* 2023;388(14):1259–1271.
6. Huo X, Ma G, Tong X, Zhang X, Pan Y, Nguyen TN, et al. Trial of Endovascular Therapy for Acute Ischemic Stroke with Large Infarct. *N Engl J Med* 2023;388(14):1272–1283.
7. Huo X. Endovascular therapy vs medical management for large core ischemic stroke: ANGEL-ASPECT results. *VJNeurology*. International Stroke Conference 2023; interview published 21 February 2023.

VALUE

Sharshar T, Porcher R, Asfar P, Grimaldi L, Jabot J, Argaud L, et al. Valproic acid as adjuvant treatment for convulsive status epilepticus: a randomised clinical trial. Crit Care 2023;27:8.

Introduction

The disappearance of convulsions is not necessarily the end of status epilepticus, nor does it guarantee neurological recovery. Patients may remain unconscious because of continuing electrographic seizures, the underlying cerebral insult, or the drugs administered to stop the seizures. For the intensive care clinician, these possibilities create different therapeutic questions: whether more treatment is needed to terminate ongoing seizures, whether recurrence can be prevented, and whether additional treatment can protect the injured brain.^{1,2}

The stepwise approach underpinning VALUE used a benzodiazepine followed, when indicated, by a second-line anti-seizure medicine. Failure could require anaesthetic infusions and prolonged critical care. The investigators considered that an adjunct with both antiseizure activity and potential neuroprotective properties might improve recovery beyond that achieved by this initial strategy. Importantly, such an approach would expose some patients whose visible seizures had already stopped to another drug, rather than reserve escalation for demonstrably persistent status.^{1,3}

Valproic acid was attractive because of its anticonvulsant activity and experimental neuroprotective rationale, including effects on inflammatory and excitatory pathways. These mechanisms provided a hypothesis, not clinical proof of neuroprotection. At the time of trial design, valproate was not part of the recommended French second-line strategy described by the investigators, making it possible to study an additional agent without routinely duplicating previous treatment. Potential adverse effects, including hyperammonaemic encephalopathy and hepatic toxicity, remained relevant when adding treatment to an already complex regimen.¹

VALUE therefore asked whether a brief intravenous course of valproic acid, added to initial antiseizure treatment in adults admitted to intensive care with generalised convulsive status epilepticus, would increase discharge alive by day 15. It did not compare valproate with another second-line agent. That distinction is fundamental: evidence against routine supplementation after initial treatment need not be evidence against the same drug when chosen to treat continuing seizures.¹

Synopsis

VALSE was an investigator-initiated, multicentre, parallel-group, double-blind, placebo-controlled randomised trial conducted in 16 French ICUs, equally divided between general and university hospitals. Recruitment ran from 18 February 2013 to 5 July 2018. Funding came from the French Ministry of Health clinical research programme, without industry involvement; an independent board monitored safety. The trial was registered as NCT01791868.^{1,2}

Adults admitted for generalised convulsive status epilepticus were eligible. This required continuous or recurrent generalised convulsions for at least five minutes without recovery of consciousness between seizures. Neither continuing visible seizures nor persistent unconsciousness was required at randomisation. The main eligibility description required antiseizure treatment to have begun within six hours before inclusion; the appendix also describes a permitted 24-hour window for persisting or recurrent status, used in 12 patients. Non-convulsive status without preceding motor manifestations and post-anoxic status were excluded, but progression from convulsive to non-convulsive status was permitted.^{1,2}

Other important exclusions were previous valproate exposure, major hepatic contraindications, an expected ICU stay below 12 hours, life expectancy below three months, and another disease expected to require hospitalisation beyond 15 days. The final eligibility description also excluded women of childbearing age between 18 and 49 years, patients under guardianship and those without health insurance. Consent could be obtained from the patient or next of kin, with deferred consent permitted when necessary.^{1,2}

Patients were allocated 1:1 through a concealed central web-based system, stratified by centre, age of 65 years or less versus older age, and acute brain injury. The intervention was intravenous valproic acid 30 mg/kg over 15 minutes, followed by 1 mg/kg/hour for 12 hours. Controls received an identical bolus and infusion of 0.9% saline. Treatment assignment was blinded, including for primary-outcome collection and EEG interpretation.¹

Both groups received a standardised initial strategy using clonazepam or diazepam, followed when indicated by phenobarbital, fosphenytoin or levetiracetam. Second-line treatment was intended to precede the study infusion. Recommended escalation included propofol or midazolam for refractory status and thiopental for super-refractory status. Maintenance medication was chosen by the treating clinician, not assigned by the trial. Supportive care included control of temperature, blood pressure, glucose, sodium and respiratory gases, together with investigation of the cause.¹

Neurological assessment occurred every four hours. A standard EEG was recommended within 24 hours and again between days 2 and 7, rather than continuous EEG monitoring. Blood sampling before treatment, after loading and after the infusion was planned to assess valproate exposure. The primary endpoint was discharge alive from hospital to home or a long-term care facility by day 15. Patients medically fit for discharge but retained for non-medical reasons could count as successes; no patient ultimately required this exception.^{1,2}

Secondary outcomes included seizure recurrence during ICU admission after initial control, development of refractory or super-refractory status, adverse events, and day-90 cognitive and functional assessment. The planned instruments were the Mini-Mental State Examination (MMSE), Frontal Assessment Battery (FAB) and extended Glasgow Outcome Scale. Mortality, awakening and lengths of stay were also collected.¹

The sample-size calculation sought a 20-percentage-point absolute increase in discharge alive, from 50% to 70%, with 90% power and a two-sided significance level of 0.05. The resulting requirement of 124 patients per group was increased to 150 per group to allow for treatment-administration errors. Analysis retained randomised allocation irrespective of actual treatment, excluding only patients who prohibited use of their data. Two missing primary outcomes were classified as failures; worst-case and available-case sensitivity analyses were specified.¹

Binary outcomes were analysed using adjusted log-binomial models with robust standard errors accounting for centre. Risk differences were derived by regression standardisation. Time-to-event analyses used Cox or competing-risk methods as appropriate. The analysis plan changed from the original logistic-model approach; its final amendment addressing non-convergent models was not made under the same treatment blinding as the earlier plan. Secondary outcomes had no multiplicity adjustment and no missing-data imputation.¹

Of 886 patients screened, 245 were randomised. One placebo recipient withdrew permission for all data use, leaving 126 valproate and 118 placebo patients in the principal analysis. The sponsor stopped recruitment because of enrolment difficulties. Seven valproate and three placebo patients did not receive their assigned treatment; another allocated to valproate received placebo. These deviations did not otherwise remove patients from the primary analysis.^{1,2}

Median age was 58.5 versus 56.2 years, and approximately two-thirds were men. About half had pre-existing epilepsy. Visible seizures had stopped in 87% of assessed patients in each group, although only about 10% had regained consciousness. Refractory status

was present in 22 patients, or 9%, at admission. Mechanical ventilation was used in 165 patients initially and a further 41 within 24 hours. Acute brain injury was recorded in 22% versus 14%, and pre-randomisation phenobarbital exposure in 23% versus 11%, in the valproate and placebo groups respectively.¹

Study infusion began a median of 35 minutes after randomisation and five hours after seizure onset. A second-line agent preceded it in 227 patients, or 93%. Valproate concentrations were available in only 22 intervention patients; the investigators reported therapeutic exposure in those tested. Maintenance-treatment data were available for 199 patients: 78/99 versus 75/100 received maintenance medication during the first 48 hours, including open-label valproate in six patients in each group. A 30-minute EEG was obtained in 168/238 patients within 24 hours and 83/201 between days 2 and 7. Persistent seizures were identified in 17 and six assessed patients, respectively.^{1,2}

Discharge alive by day 15 occurred in 77/126 valproate patients and 72/118 controls, both 61%. Using the estimates in Table 2, the adjusted relative risk was 1.04 (95% CI 0.90–1.20; P=0.58), with an adjusted risk difference of 2.5 percentage points (95% CI –6.3 to 11.2). Both sensitivity analyses gave similar results. Refractory status developed in 5/126 versus 6/118 patients and super-refractory status in 3/126 versus 4/118, without statistically significant differences.¹

ICU seizure recurrence occurred in 7/125 versus 14/118 patients, or 6% versus 12% (adjusted relative risk 0.48; 95% CI 0.23–1.00; P=0.049). Median awakening time was 24 versus 18 hours, ICU stay five versus four days, and hospital stay 12 versus 11 days; none differed significantly. Hospital deaths numbered seven versus two. By day 90, eight versus three deaths were recorded (adjusted hazard ratio 2.32; 95% CI 0.61–8.84; P=0.22). Day-90 MMSE, FAB and Glasgow outcome assessments were available for only 39, 49 and 99 patients, respectively, without demonstrated between-group improvement.¹

Using Table 3, at least one adverse event affected 45/126 valproate patients versus 52/118 controls, or 36% versus 44% (P=0.19). Serious adverse events affected 33/126 versus 27/118, or 26% versus 23% (P=0.56). Hepatic cytolysis was reported in four versus three patients. The appendix separately records 24 seizure-recurrence adverse events, 21 after ICU discharge; these were not the same outcome or counting unit as patients with recurrence during ICU admission.^{1,2}

Critique

The principal result is evidence against routine adjunctive valproate in the population and treatment sequence studied, not a verdict on valproate for every form of status epilepticus. Most patients had already stopped visibly convulsing and almost all received a second-line agent before the trial infusion. VALSE therefore largely tested whether an additional short course could improve the subsequent trajectory, rather than whether valproate could terminate benzodiazepine-resistant seizures as the selected second-line treatment.¹

Several design features support confidence in this comparison. Central concealed allocation and an identical placebo reduced opportunities for treatment-selection and performance bias. Primary-outcome assessors were blinded, and hospital status was known for all but two of the 244 analysed participants. Retaining patients who did not receive the assigned drug preserved the randomised comparison. The agreement of missing-data sensitivity analyses makes those two missing outcomes an unlikely explanation for the null primary finding.¹

The population nevertheless limits what can be inferred about resistant seizures. Only 9% were classified as refractory at admission, while some patients were already awake. Conversely, the high rate of ventilation does not prove persistent epileptic activity: airway management can reflect sedative exposure or the underlying illness. Including clinically controlled patients was consistent with the proposed prevention and neuroprotection strategy, rather than necessarily a design error. It does mean that a targeted trial in EEG-confirmed refractory status could answer a different question.^{1,2}

Discharge alive is clinically relevant but an indirect measure of the proposed drug effects. It combines survival with the speed of recovery and the ability to leave acute hospital care. A transfer to long-term care qualifies alongside return home, so success does not establish independence or preserved cognition. Underlying brain disease, investigations and maintenance-treatment decisions may influence discharge without being altered by a 12-hour infusion. The non-medical-delay exception addressed one potential problem, but was not used. The null result cannot establish that seizure control itself is unimportant to recovery.^{1,2}

Early termination should be interpreted through precision, not simply the missed recruitment target. The trial fell short of 300 patients but enrolled almost the uninflated requirement of 248. Its adjusted confidence interval excludes the anticipated 20-percentage-point improvement, while allowing benefits up to 11.2 points and worsening up to 6.3 points. Thus, the large planned effect was not supported, but a smaller clinically important benefit was not excluded. Identical rounded event

percentages also do not imply exact equivalence; the adjusted effect estimates account for prognostic variables and differ from the crude comparison.¹

Treatment exposure raises narrower uncertainties. A median five-hour interval from seizure onset is not a test of immediate adjunctive administration. Therapeutic concentrations in 22 patients support delivery in that subgroup, not verified adequate exposure throughout the whole cohort or proof of neuroprotective target engagement. Maintenance treatment was discretionary and included valproate in both groups. This does not erase the randomised contrast, but confines it to adding the specified early course to a subsequently evolving regimen. Earlier, longer or higher-dose treatment might behave differently; VALSE provides no evidence that it would work better.^{1,2}

Limited EEG surveillance weakens interpretation of the seizure outcomes. A short recording in 71% of patients during the first day, and fewer thereafter, cannot reliably quantify total electrographic seizure burden. Clinical improvement influenced whether recordings were obtained, alongside availability, so the recorded subgroup was selected rather than representative by design. Some persistent activity could have been missed, while impaired consciousness could have been attributed to seizures when another explanation applied. These uncertainties limit mechanistic conclusions, although they do not directly invalidate the observed hospital-discharge endpoint.^{1,2}

The recurrence finding deserves attention without promotion to a definitive benefit. Its P value of 0.049 arose among multiple unadjusted secondary comparisons, with only 21 affected patients. Table 2 also gives an adjusted absolute difference of –6.1 percentage points with a 95% interval from –15.2 to 3.0. This imprecision and the absence of improvements in refractory status, awakening or discharge argue against using recurrence alone to justify routine treatment. Furthermore, adverse-event recurrences occurring after ICU discharge cannot be substituted for the prespecified ICU recurrence outcome; event counts include repeated episodes in individual patients.^{1,2}

The neuroprotection hypothesis was especially poorly resolved. MMSE and FAB results were obtained in only 16% and 20% of the cohort, and functional assessments in about 41%. The ability and willingness to attend follow-up may relate to recovery, making available-case comparisons vulnerable to selection. Similar scores among the few assessed cannot establish preserved cognition across the randomised population. In contrast to the nearly complete primary endpoint, these long-term data provide little reassurance about an effect that might not shorten hospitalisation.^{1,2}

Safety should also remain qualified. No statistically clear excess in reported adverse events was demonstrated, but serious events were not less frequent with valproate and

the mortality estimate was compatible with considerable harm as well as benefit. The small numbers and competing underlying causes prevent attributing the numerical mortality difference to treatment. Conversely, a non-significant mortality comparison does not establish safety. Major hepatic contraindications were excluded, exposure was brief, and laboratory testing followed clinical prescription rather than a documented universal surveillance schedule for every potential toxicity.^{1,2}

Finally, the published protocol appeared in 2021, after recruitment had ended, although it stated that treatment remained blinded. The final report acknowledges analysis-plan changes and a last amendment outside the earlier blinding arrangements. The protocol also lists a childbearing-age exclusion amendment approved after the reported recruitment end, making its timing in practice unclear. Clarification is needed before assuming either selective analysis or that the final eligibility rules applied unchanged throughout recruitment. Overall, VALSE argues against automatic supplementation, while leaving targeted rescue treatment, smaller effects and long-term neuroprotection unresolved.^{1,3}

Body of Evidence

The relevant evidence addresses several different decisions along the treatment pathway. Initial seizure termination, selection of a second-line agent, addition of another medicine regardless of response, and prevention of subsequent disability are not interchangeable questions. Comparing studies requires attention to the stage at which patients were enrolled and to whether the outcome measured visible seizure cessation, recovery of consciousness or a later patient-centred result.^{1,4}

RAMPART demonstrated the importance of a deliverable first-line strategy. In this double-blind pre-hospital trial, intramuscular midazolam was compared with intravenous lorazepam in children and adults who remained convulsing after more than five minutes. Seizures had stopped without rescue treatment by emergency-department arrival in 329/448 versus 282/445 patients, or 73.4% versus 63.4%. The difference was 10 percentage points (95% CI 4.0–16.1), satisfying both non-inferiority and superiority criteria. Among patients who responded before arrival, active treatment was delivered sooner intramuscularly, despite faster subsequent seizure cessation after intravenous dosing. This was evidence for improving the initial delivery strategy, not for adding a later prophylactic agent after that strategy had succeeded.⁵

SAMUKeppra tested an earlier adjunctive approach more directly. Adults receiving pre-hospital clonazepam were assigned additional levetiracetam or placebo in a randomised double-blind comparison. Among the 136 patients in the primary analysis,



convulsions stopped within 15 minutes in 50/68 patients receiving levetiracetam and 57/68 receiving placebo, or 74% versus 84% (difference –10.3 percentage points; 95% CI –24.0 to 3.4). The addition did not improve early clinical seizure control. This does not predict the effect of every combination, but cautions against assuming that another active antiseizure drug must improve outcomes when an effective initial treatment is already being administered. VALSE extended the adjunctive question into the ICU, after additional treatment and with discharge rather than immediate convulsion cessation as its primary outcome.^{1,6}

The VALSE investigators also placed their neuroprotective rationale alongside HYBERNATUS, which evaluated adjunctive hypothermia in ICU patients with convulsive status epilepticus. Their account describes no improvement in neurological outcome and an association with serious adverse events. The comparison is conceptually relevant: an intervention intended to modify secondary injury requires evidence of better patient recovery, not merely a plausible mechanism. It is not a direct comparison with valproate, and the differing populations and surveillance methods prevent treating the two trials as replications of the same strategy.^{1,2}

ESETT addressed the distinct second-line choice. This blinded adaptive randomised trial enrolled 384 children and adults whose convulsive status had not responded to benzodiazepines: 145 received levetiracetam, 118 fosphenytoin and 121 valproate. The primary endpoint required cessation of clinically apparent seizures and improved consciousness at 60 minutes without further anticonvulsant treatment. It occurred in 47%, 45% and 46%, respectively; the corresponding 95% credible intervals were 39–55%, 36–54% and 38–55%. Recruitment stopped under a prespecified rule for futility of identifying a superior or inferior drug. Approximately half responded to each strategy, without a clear safety difference. This was neither proof of exact equivalence nor evidence that all three should be administered together. It supports separating valproate as one second-line option from valproate added after another second-line agent.⁴

These studies therefore inform different parts of care rather than compete for a single preferred regimen. RAMPART concerns prompt initial delivery; SAMUKeppra and VALSE question routine additional treatment; ESETT compares alternatives after benzodiazepine failure. VALSE adds a blinded assessment of a brief adjunctive valproate course in a predominantly clinically controlled ICU population. Its negative discharge result should discourage automatic escalation without an ongoing indication, but should not be extended to prohibit valproate when selected for a different, evidence-supported treatment role.^{1,4,5,6}

Should we routinely add valproic acid after first- and second-line treatment of convulsive status epilepticus?

No. VALSE did not improve discharge alive by day 15; it does not challenge valproate as a second-line treatment.

References

1. Sharshar T, Porcher R, Asfar P, Grimaldi L, Jabot J, Argaud L, et al. Valproic acid as adjuvant treatment for convulsive status epilepticus: a randomised clinical trial. *Crit Care* 2023;27:8.
2. Sharshar T, Porcher R, Asfar P, et al. Supplementary appendix to: Valproic acid as adjuvant treatment for convulsive status epilepticus: a randomised clinical trial. *Crit Care* 2023;27:8.
3. Sharshar T, Ben Hadj Salem O, Porcher R, Grimaldi-Bensouda L, Heming N, Clair B, et al. Valproic Acid as an Adjuvant Treatment for Generalized Convulsive Status Epilepticus in Adults Admitted to Intensive Care Units: Protocol for a Double-Blind, Multicenter Randomized Controlled Trial. *JMIR Res Protoc* 2021;10(2):e22511.
4. Kapur J, Elm J, Chamberlain JM, Barsan W, Cloyd J, Lowenstein D, et al. Randomized Trial of Three Anticonvulsant Medications for Status Epilepticus. *N Engl J Med* 2019;381(22):2103–2113.
5. Silbergleit R, Durkalski V, Lowenstein D, Conwit R, Pancioli A, Palesch Y, et al. Intramuscular versus intravenous therapy for prehospital status epilepticus. *N Engl J Med* 2012;366(7):591–600.
6. Navarro V, Dagrón C, Elie C, Lamhaut L, Demeret S, Urien S, et al. Prehospital treatment with levetiracetam plus clonazepam or placebo plus clonazepam in status epilepticus (SAMUKeppra): a randomised, double-blind, phase 3 trial. *Lancet Neurol* 2016;15(1):47–55.

Possible Reporting Inconsistencies. The adverse-event paragraph appears to reverse the treatment labels. Table 3 reports at least one adverse event in 45/126 valproate recipients and 52/118 placebo recipients; the prose assigns the counts to the opposite treatments. The table is used in this chapter.¹

DASH

Desborough MJR, Al-Shahi Salman R, Stanworth SJ, Havard D, Woodhouse LJ, Craig J, et al.; DASH trial investigators. Desmopressin for patients with spontaneous intracerebral haemorrhage taking antiplatelet drugs (DASH): a UK-based, phase 2, randomised, placebo-controlled, multicentre feasibility trial. Lancet Neurol. 2023;22(7):557–567.

Introduction

An intracerebral haemorrhage in a patient taking antiplatelet medication creates an immediate therapeutic dilemma. The medication may impair haemostasis, yet correcting a laboratory abnormality does not necessarily improve neurological recovery. Treatment must act while bleeding remains modifiable and avoid introducing thrombosis, fluid retention or other complications. A plausible haemostatic mechanism is therefore a starting point for investigation, rather than sufficient evidence to treat every exposed patient.¹

Desmopressin offers a comparatively simple intervention: a short intravenous infusion intended to augment haemostasis through increased von Willebrand factor availability and related effects. A small prospective study demonstrated improved platelet-function measurements after treatment in patients with intracerebral haemorrhage. That observation supported biological plausibility but could not determine whether desmopressin reduced disability or death. The relevant clinical question remained whether an additional dose improved outcomes when given alongside standard stroke care.²

Experience with platelet transfusion reinforced the need for randomised evidence. PATCH found worse functional outcomes with platelet transfusion in patients with antiplatelet-associated intracerebral haemorrhage. This did not establish that desmopressin would be harmful, because the treatments differ, but it showed why an apparent need to reverse antiplatelet effects cannot substitute for clinical outcome testing. The feasibility of conducting such testing also mattered: patients may lack capacity, arrive outside research hours and need treatment before conventional trial processes can be completed.³

DASH was designed to resolve those practical uncertainties before a definitive efficacy trial. It tested whether eligible patients could be identified, randomised and treated, whether follow-up was achievable, and what recruitment rate a larger study might

expect. Clinical and laboratory outcomes were collected to inform that next step. The distinction is fundamental: DASH was not designed to establish that desmopressin prevents death or dependence, even if its small clinical comparisons appeared encouraging.

Synopsis

DASH was a phase 2, randomised, placebo-controlled feasibility trial at ten UK hospitals recruiting through emergency departments and acute stroke services. Recruitment occurred from 1 April 2019 to 31 March 2022. The National Institute for Health Research funded the study and had no role in its design, conduct, analysis, interpretation or writing. Recruitment was interrupted by the COVID-19 pandemic, and the study period was extended. The published report describes amendments and their rationale; it states that the statistical analysis plan was agreed before database lock and release of allocation codes.¹

Eligible adults had CT-confirmed spontaneous intracerebral haemorrhage and were prescribed, and thought to have taken, a daily oral antiplatelet drug during the preceding seven days. The permitted interval from symptom onset to randomisation was initially 12 hours. Because recruitment was slow, an amendment dated 22 November 2019 extended this to 24 hours. Exclusions included a known secondary cause of haemorrhage, risk of fluid retention, systolic pressure below 90 mmHg, a drug-eluting stent within three months, desmopressin allergy, pregnancy or breastfeeding, anticipated survival below four hours or palliative-only care, Glasgow Coma Scale score below five and premorbid modified Rankin Scale score above four.¹

Patients with capacity consented directly. Otherwise, a relative or close friend could act as representative, or an independent doctor could provide proxy consent when a personal representative was unavailable. Patients were approached again if capacity returned. Central computer-based allocation used a 1:1 ratio, minimisation and a 5% random flip. Factors included age at least 70 years, sex, onset at least three hours earlier, systolic pressure at least 170 mmHg and intraventricular haemorrhage status. Clinicians, patients and clinical, radiological and laboratory outcome assessors were masked. Personnel preparing injections could not be masked because the commercial ampoules differed.¹

Desmopressin was administered as a single fixed intravenous dose of 20 micrograms, prepared from five 4-microgram ampoules added to 50 mL of 0.9% sodium chloride and infused over 20 minutes. Placebo consisted of saline ampoules added to a 50-mL saline bag and administered over the same duration. The fixed dose was chosen for practicality

and to reduce emergency dosing errors; it was not an individual weight-based regimen. Both arms also received usual care, including blood-pressure management, intermittent pneumatic compression and neurosurgery when appropriate. Sodium was checked locally at 24 hours and made available to the treating team.¹

The primary outcomes concerned feasibility: eligibility and recruitment proportions, reasons for non-enrolment, recruitment per site-month, time to randomisation, delivery of allocated treatment, follow-up and completeness of outcome collection. No formal efficacy sample-size calculation was performed. The investigators considered recruitment of more than 50 participants from ten centres in one year a useful basis for planning a larger trial. Independent data monitoring reviews occurred during recruitment. The study therefore had neither a treatment-effect target nor statistical power intended to establish a clinical benefit.¹

Secondary outcomes included haematoma growth at 24 hours, sodium disturbance, hospital stay, discharge destination, mortality, adverse events and 90-day function, quality of life, cognition and mood. Centrally assessed haematoma expansion meant an increase above 6 mL or greater than 33%. The designated death-or-dependency outcome used modified Rankin Scale scores above four, meaning scores five or six. The full ordinal scale was also analysed. Laboratory work assessed von Willebrand factor antigen and activity and factor VIII at baseline and one hour, with P-selectin testing intended to characterise baseline antiplatelet effects.¹

Clinical summaries followed allocation; safety summaries used treatment received. All available observations were used without imputation. The two Rankin comparisons used unadjusted ordinal and binary logistic regression with 95% confidence intervals. No formal comparisons were undertaken for the other secondary outcomes. General adverse events were collected during the first 24 hours; serious-event assessment and specified safety follow-up continued over longer windows, with fatal events and safety outcomes followed to day 90. Central assessors conducted masked 90-day telephone follow-up.¹

Of 1,380 screened patients, 176 were potentially eligible, 57 were approached and 54 were randomised, 27 per group. Among 122 eligible patients not recruited, 74 arrived outside working hours, 17 lacked a doctor available for consent and ten encountered recruitment on hold. Twenty-three participants consented personally and 31 through representatives. Before widening the window, 14 were enrolled over eight months; 11 more were enrolled over four months before the pandemic, followed by 29 over 22 months. The reported mean recruitment rate was 0.2 participants per centre-month, with a 95% CI of 0.1–0.3.^{1,4}

Participants were predominantly older adults, 36/54 were male and 50/54 White. Median Glasgow Coma Scale score was 15 in each arm. Median baseline haematoma volume was 13.3 mL with desmopressin and 8.1 mL with placebo, among those with measurements. Aspirin alone accounted for 13 versus 12 participants and clopidogrel alone for 14 versus nine. All six participants receiving dual antiplatelet therapy were allocated to placebo. Median onset-to-treatment time was 10.8 versus 10.4 hours, and randomisation-to-treatment time was 0.6 versus 0.8 hours. Among 40 participants enrolled after the window amendment, 21 were randomised at least 12 hours after onset.¹

Twenty-six of 27 desmopressin patients and all 27 placebo patients received the full infusion. A leaking saline bag caused the one partial dose. There were no withdrawals, crossovers or participants lost to follow-up, and all 54 had a 90-day Rankin outcome. However, completeness varied for other measurements. Drug storage created an additional operational burden: 100/140 treatment packs experienced at least one temperature excursion, 69 were quarantined and 59 destroyed. Twenty-five participants received a pack that had experienced an excursion; this was not itself evidence that they received an ineffective or unsafe preparation.¹

At day 90, death or dependency as defined by Rankin scores five or six occurred in 6/27 desmopressin patients versus 10/27 placebo patients, unadjusted odds ratio 0.49 (95% CI 0.15–1.61). The full-scale ordinal odds ratio was 0.74 (95% CI 0.29–1.93). Deaths numbered five versus seven at 90 days. Centrally assessed haematoma expansion occurred in 5/24 versus 5/22 patients, with median volume changes of zero versus 0.5 mL. These were exploratory clinical observations within a feasibility study.¹

At 24 hours, sodium below 135 mmol/L occurred in 8/23 versus 3/24 participants; none of the 47 measured values was below 125 mmol/L. Serious adverse events affected 12/27 versus 13/27 participants, comprising 16 versus 22 events. One thrombotic event occurred in each arm: myocardial infarction with desmopressin and ischaemic stroke with placebo. Haemostatic markers were available in only 24 participants and showed descriptively greater increases after desmopressin without formal comparison. Baseline P-selectin testing was completed in 14 patients; only three of nine aspirin exposures and four of six clopidogrel exposures met the assay's inhibition thresholds. One participant received both agents. The test supplier ceased trading, preventing completion and further validation.^{1,4}

Critique

DASH demonstrates that patients can be randomised to desmopressin or placebo, receive treatment reliably and complete functional follow-up. It does not demonstrate that a definitive trial would recruit easily at the originally envisaged rate. The practical success after enrolment and the difficulty reaching enrolment are separate findings. Fifty-four participants were recruited over three years, rather than more than 50 in one year as initially envisaged. Pandemic disruption clearly contributed, but slow recruitment had already led to a wider eligibility window. The case for proceeding therefore depends on changing delivery arrangements, not merely reproducing the feasibility study on a larger scale.^{1,4}

The screening data identify an actionable problem. Only three eligible patients declined, whereas out-of-hours presentation and lack of a consenting doctor accounted for most missed opportunities. This supports integration with ordinary emergency stroke services, where clinical staff can recruit without waiting for a separate research team. It does not mean that all missed patients would necessarily have consented or been treatable within an effective biological window. Screening logs were recorded locally and could not be centrally verified, and the participating hospitals were research-active stroke centres. Their apparent opportunities should inform planning without being treated as guaranteed recruitment.^{1,4}

Widening the window also changes the therapeutic question. It improves access to enrolment while admitting patients after much of the early bleeding evolution may already have occurred. Median treatment was more than ten hours after onset, and over half of those enrolled after the amendment were randomised at least 12 hours after onset. DASH therefore cannot tell whether earlier desmopressin is effective, whether treatment remains useful late, or whether a broad 24-hour window is optimal. A definitive trial must balance achievable recruitment against the chance that the intervention arrives while its target remains modifiable. Feasibility should not be improved at the cost of making a clinically relevant effect harder to detect.¹

The treatment-delay data expose another opportunity. Baseline imaging occurred well before randomisation for many patients, and the median interval from scanning to treatment exceeded four hours overall. Delay was not confined to late arrival. Streamlining consent, access to treatment packs and coordination might both increase recruitment and bring treatment earlier. The trial did not randomise a streamlined pathway, so its benefit remains a proposal rather than an observed effect. Nonetheless, these process measurements are precisely the information a feasibility study should provide before committing to a large clinical comparison.¹

The clinical estimates must remain subordinate to that purpose. The binary odds ratio favours desmopressin but its interval ranges from a large benefit to appreciable harm. The ordinal interval is similarly broad. Neither finding establishes benefit, equivalence or futility. Sixteen events in the binary comparison provide little protection against unstable estimates and chance baseline imbalance. All dual-antiplatelet patients entered the placebo arm, while haematoma volume and some other prognostic characteristics differed in the opposite direction. It is impossible to determine how these imbalances account for the outcome pattern, and a post-hoc adjustment in such a small sample would not create the information of a larger trial.¹

The binary outcome's label also needs precision. Scores five and six represent severe disability or death; score four already includes substantial dependence but is outside this dichotomy. Calling the result freedom from all dependence would overstate it. The ordinal analysis retains information across the scale and is therefore a useful companion, although its uncertainty remains considerable. Complete Rankin follow-up is a genuine strength. It should not be broadened into a claim that every outcome was complete: repeat imaging was available in 46 participants, sodium in 47, and cognitive and mood assessments in substantially fewer.¹

The biological observations do not establish successful reversal of an antiplatelet defect in every enrolled patient. Eligibility depended on a medication history rather than a validated demonstration of inhibited platelets. In the small tested subset, inhibition was frequently absent by the chosen thresholds. This could reflect adherence, variable drug response or assay performance; it cannot establish widespread resistance or non-adherence. Loss of the assay supplier compounded the uncertainty. The result is a pragmatic comparison in people believed to be taking antiplatelet drugs, not a confirmed biological subgroup whose dysfunction was shown to normalise.^{1,4}

Similarly, the haemostatic-marker substudy was small and imbalanced at baseline. Descriptive increases after treatment support the intended direction of biological activity but cannot explain the clinical estimates. The fixed 20-microgram dose improves simplicity, yet DASH did not compare it with weight-based dosing, alternative doses or repeat administration. A decision that is practical for a large trial may be sensible without being the proven optimum. Neither the absence of a clear imaging difference nor the favourable functional point estimate identifies which dose or patient subgroup should be selected.¹

Safety remains incompletely resolved. More measured patients had mild hyponatraemia after desmopressin, making sodium monitoring an important component of subsequent research. No observed value below 125 mmol/L is reassuring only within the small



measured sample and this trial's definition. One thrombotic event in each arm cannot exclude a clinically important increase in thrombotic risk, especially when recent coronary stenting and fluid-retention risk were exclusions. The discussion's broad reassurance should not replace the actual event counts. Serious-event categories and imaging-defined haematoma growth also differ: four versus six serious events labelled haemorrhagic expansion are not the same comparison as five versus five centrally measured expansions.¹

Masking, central randomisation and blinded outcome assessment protect the comparison. The preparation staff exception is clearly relevant, however, and the study does not establish that treatment was indistinguishable at every step. Temperature excursions chiefly demonstrate a burdensome supply process; without potency measurements they should not be assumed to explain the results. The conflicting 28-day mortality counts noted below do not prevent appraisal of the consistent primary feasibility counts or 90-day Rankin results. They do require care when interpreting early mortality.¹

Body of Evidence

Naidech and colleagues' 2014 prospective study treated 14 patients with intracerebral haemorrhage and aspirin exposure, reduced platelet activity or both, using desmopressin 0.4 micrograms/kg. Platelet-function analyser closure time improved one hour after treatment and von Willebrand factor antigen increased. Without a randomised comparator and with so few patients, the study established a biological signal rather than a clinical outcome benefit. DASH subsequently moved the question into a masked randomised design, but used a fixed dose and was still directed primarily at feasibility.²

PATCH supplies a different but essential comparison. It randomised 190 patients with recent antiplatelet exposure and acute supratentorial intracerebral haemorrhage to platelet transfusion or standard care. The adjusted common odds ratio for a worse three-month outcome was 2.05 (95% CI 1.18–3.56), favouring standard care. This finding is specific to platelet transfusion and its trial population; it cannot be transferred as an estimate of desmopressin's effect. Its broader lesson is that a strategy intended to restore haemostasis can worsen patient outcomes and therefore requires direct testing.³

TICH-2 tested another haemostatic approach, tranexamic acid, in 2,325 patients treated within eight hours of intracerebral haemorrhage onset. Functional status at 90 days did not improve significantly: adjusted ordinal odds ratio 0.88 (95% CI 0.76–1.03). Fewer early deaths did not translate into a clear 90-day mortality difference. This larger trial

does not answer the antiplatelet-specific desmopressin question, but shows why an early outcome signal cannot replace later functional assessment. DASH's emphasis on complete 90-day Rankin follow-up is consequently an important preparation for definitive testing.⁵

A multicentre retrospective study by Summers and colleagues, published in 2023, included 209 patients with antiplatelet-associated intracranial haemorrhage, 118 treated with desmopressin and 91 untreated. Haematoma expansion occurred in 16.1% versus 17.6%, without a significant difference. Its expansion definition was greater than 20% growth, differing from DASH, and eligibility required follow-up imaging and excluded some short admissions and intervening major surgery. These data do not establish equivalence. Treatment selection and retrospective inclusion can influence observed outcomes, while the broader intracranial-haemorrhage population is not automatically identical to spontaneous intracerebral haemorrhage.⁶

The evidence available by the end of 2023 thus separated three questions: whether desmopressin changes haemostatic measurements, whether emergency randomisation is possible, and whether patients ultimately do better. The first had a small mechanistic signal; DASH made progress on the second; the third remained unanswered. The studies cannot be combined conceptually as if favourable laboratory results, achievable follow-up and an imprecise clinical odds ratio together constituted proof of benefit. Each step addresses a different uncertainty.

DASH's most useful contribution is a realistic account of the next trial's requirements. Recruitment must be available outside research hours; treatment should follow diagnosis promptly; masking and sodium surveillance must remain credible; and outcome collection must include function, mortality and harms. Decisions about the time window, medication verification and dose require an explicit clinical rationale rather than an assumption that the feasibility configuration is optimal. A larger study could reasonably proceed on that basis, but routine efficacy claims would have preceded the evidence. DASH supports completing the randomised evaluation of desmopressin, while preserving uncertainty about its clinical value.

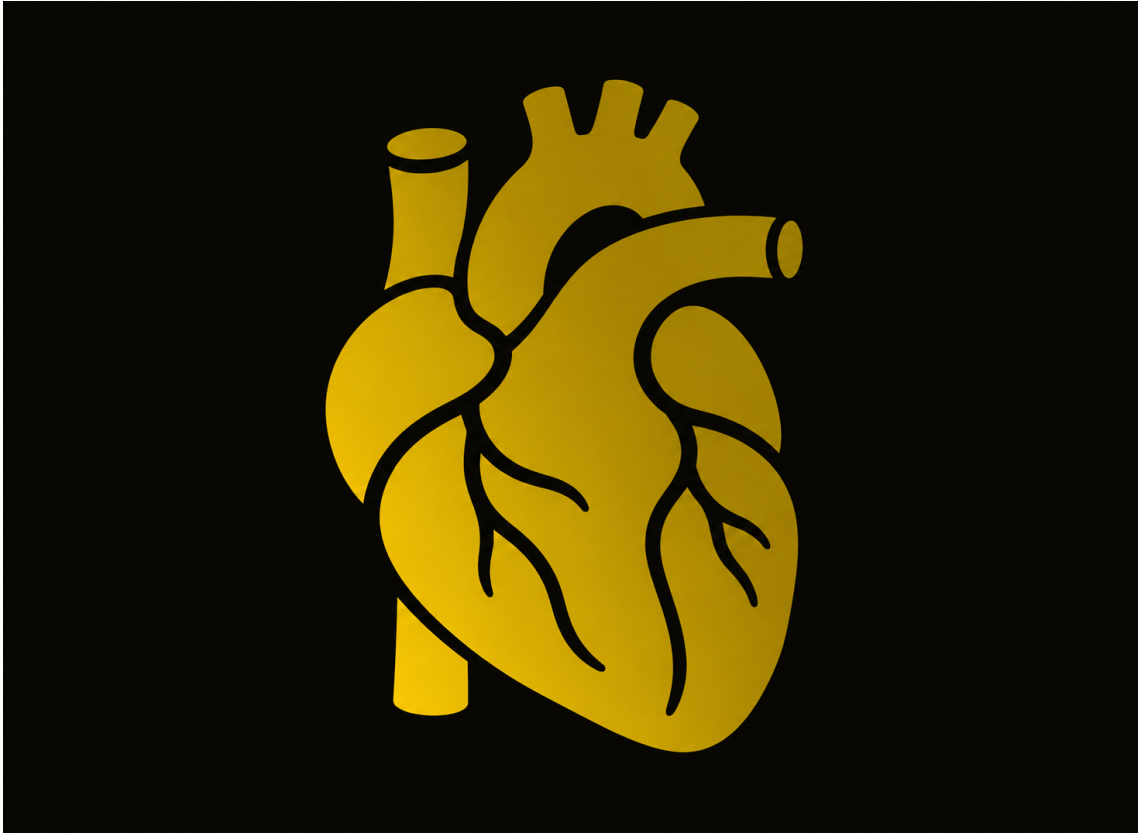
Does DASH establish that desmopressin improves antiplatelet-associated intracerebral haemorrhage?

No. DASH supports a definitive randomised trial, with better recruitment pathways; it does not establish clinical efficacy or exclude important harm.

References

1. Desborough MJR, Al-Shahi Salman R, Stanworth SJ, Havard D, Woodhouse LJ, Craig J, et al.; DASH trial investigators. Desmopressin for patients with spontaneous intracerebral haemorrhage taking antiplatelet drugs (DASH): a UK-based, phase 2, randomised, placebo-controlled, multicentre feasibility trial. *Lancet Neurol.* 2023;22(7):557–567.
2. Naidech AM, Maas MB, Levasseur-Franklin KE, Liotta EM, Guth JC, Berman M, et al. Desmopressin improves platelet activity in acute intracerebral hemorrhage. *Stroke.* 2014;45(8):2451–2453.
3. Baharoglu MI, Cordonnier C, Al-Shahi Salman R, de Gans K, Koopman MM, Brand A, et al.; PATCH Investigators. Platelet transfusion versus standard care after acute stroke due to spontaneous cerebral haemorrhage associated with antiplatelet therapy (PATCH): a randomised, open-label, phase 3 trial. *Lancet.* 2016;387(10038):2605–2613.
4. Desborough MJR, Al-Shahi Salman R, Stanworth SJ, et al. Supplementary appendix to: Desmopressin for patients with spontaneous intracerebral haemorrhage taking antiplatelet drugs (DASH): a UK-based, phase 2, randomised, placebo-controlled, multicentre feasibility trial. *Lancet Neurol.* 2023.
5. Sprigg N, Flaherty K, Appleton JP, Al-Shahi Salman R, Bereczki D, Beridze M, et al.; TICH-2 Investigators. Tranexamic acid for hyperacute primary IntraCerebral Haemorrhage (TICH-2): an international randomised, placebo-controlled, phase 3 superiority trial. *Lancet.* 2018;391(10135):2107–2115.
6. Summers A, Singh J, Lai M, Schomer KJ, Martin R, Vitt JR, et al. A multicenter retrospective study evaluating the impact of desmopressin on hematoma expansion in patients with antiplatelet-associated intracranial hemorrhage. *Thromb Res.* 2023;222:96–101.

Possible Reporting Inconsistencies. The flow-diagram footnote records six placebo-group deaths before day 28, but Table 2 and the Results give five. Both accounts give seven deaths by day 90. The timing of one placebo death therefore appears inconsistent, leaving the exact 28-day mortality count uncertain.¹



Circulatory Trials

MINT

Carson JL, Brooks MM, Hébert PC, Goodman SG, Bertolet M, Glynn SA, et al. Restrictive or Liberal Transfusion Strategy in Myocardial Infarction and Anemia. N Engl J Med 2023;389(26):2446–2456.

Introduction

Restrictive red-cell transfusion has become an established approach in many clinical settings, reducing exposure to donor blood without a demonstrated deterioration in major outcomes. Acute myocardial infarction remains a particularly important boundary to that evidence. Anaemia is common in these patients, but an association between lower haemoglobin and worse outcomes does not establish that transfusion reverses the risk. Conversely, evidence supporting restriction in other populations cannot simply establish its safety during acute myocardial ischaemia.¹

There are credible arguments in both directions. Increasing haemoglobin may improve oxygen delivery to myocardium whose blood supply is already compromised. This may be particularly relevant when coronary flow cannot increase sufficiently to compensate for anaemia. However, transfusion also introduces volume, increases blood viscosity and may have inflammatory, thrombotic or immunological effects. Patients with myocardial infarction frequently have heart failure or renal dysfunction, making the balance between oxygen delivery and circulatory tolerance especially pertinent. A higher haemoglobin concentration is therefore a plausible treatment target, not a sufficient demonstration of clinical benefit.^{1,2}

The practical uncertainty concerns more than a single laboratory threshold. Should patients be transfused proactively to maintain haemoglobin around 10 g/dL, or should transfusion be deferred until the concentration falls below 7–8 g/dL or persistent ischaemic symptoms demand intervention? Earlier small trials in coronary disease had produced conflicting signals. Evidence in postoperative patients with cardiovascular disease or risk factors was reassuring about restriction, but did not directly answer the question in patients experiencing an acute infarction.^{1,3}

The Myocardial Ischemia and Transfusion trial, MINT, was designed to provide a substantially larger randomised comparison. It tested whether a liberal transfusion strategy reduced death or recurrent myocardial infarction within 30 days in patients admitted with acute myocardial infarction and haemoglobin below 10 g/dL.¹

Synopsis

MINT was a phase 3, multicentre, open-label, randomised superiority trial conducted at 144 sites in the United States, Canada, France, Brazil, New Zealand and Australia. Recruitment ran from April 2017 to April 2023. The National Heart, Lung, and Blood Institute provided the principal funding and was represented in trial oversight; additional support included Canadian Blood Services and the Canadian Institutes of Health Research. An independent data and safety monitoring committee reviewed safety and interim efficacy data. Patients or their surrogates provided written informed consent.¹

Adults with ST-segment elevation or non-ST-segment elevation myocardial infarction, diagnosed according to the Third Universal Definition, were eligible if haemoglobin was below 10 g/dL within 24 hours before randomisation. Infarction types 1, 2, 4b and 4c were included, with diagnosis and classification undertaken locally. The attending physician had to consider both strategies appropriate. Important exclusions were uncontrolled acute bleeding, planned cardiac surgery during the admission, palliative treatment alone, refusal of transfusion and inability to obtain 30-day follow-up. Previous transfusion did not prevent enrolment. The protocol defined uncontrolled bleeding by the need for uncrossmatched or non-type-specific blood, rather than excluding every patient with ongoing blood loss.^{1,2}

Allocation was 1:1 through a central web-based system, using variable blocks of four and six and stratification by site. An independent statistician generated the sequence. Patients and treating clinicians knew the assignment. In the restrictive group, transfusion was permitted, but not required, below 8 g/dL and strongly recommended below 7 g/dL or for angina unresponsive to medical treatment. Blood was given one unit at a time, with reassessment and sufficient transfusion to address the low haemoglobin or uncontrolled symptoms.^{1,2}

In the liberal group, one unit was administered after randomisation, followed by further transfusion whenever necessary to maintain haemoglobin at or above 10 g/dL. The strategy continued until index-hospital discharge or 30 days. Haemoglobin was rechecked after each unit. Both groups could receive emergency transfusion for active bleeding, with the protocol paused when required. Transfusion could be delayed until adequate diuresis in volume-overloaded patients or coordinated with dialysis. The protocol specified leucoreduced red cells; storage characteristics, other blood products, revascularisation and other treatments were not assigned by the trial.^{1,2}

Haemoglobin, troponin and electrocardiography were assessed before randomisation and during the first three subsequent days, including two troponin measurements on the first day. Clinical events were identified during admission and through 30-day contact

and record review. The primary outcome was death from any cause or recurrent myocardial infarction within 30 days. A blinded central committee screened recorded troponin results and adjudicated suspected recurrent infarctions, including events reported by sites. Recurrent infarction was the only centrally adjudicated outcome. Cause of death was classified locally as cardiac, non-cardiac or undetermined.^{1,4}

The prespecified secondary outcomes were death, recurrent infarction, and an expanded composite of death, infarction, ischaemia-driven unscheduled coronary revascularisation or hospital readmission for an ischaemic cardiac condition, all through 30 days. Other outcomes included heart failure, thromboembolism, infection, bleeding and transfusion complications. Six-month vital status was also collected, but those results were not presented in the supplied primary report.^{1,4}

The planned sample of 3,500 provided 80% power to identify a 20% relative difference, with a two-sided significance level of 0.05. The protocol assumed primary-event rates of 18.2% with restriction and 14.6% with liberal transfusion, a pooled incidence of 16.4%, and allowed for missing outcomes. The primary intention-to-treat analysis used log-binomial regression with treatment as a fixed effect and site as a random effect. Multiple imputation addressed incomplete follow-up. Risk ratios compared restrictive with liberal treatment, so values above one favoured the liberal strategy. Secondary outcomes and subgroup analyses were not adjusted for multiplicity.^{1,2}

Of 3,530 patients who consented, 24 were ineligible and 3,506 were randomised. Two liberal-group patients did not permit use of their data, leaving 1,749 restrictive and 1,755 liberal participants for analysis. Follow-up was complete in 3,447; 20 restrictive and 37 liberal participants had incomplete primary-outcome ascertainment. Mean age was 72.1 years, 45.5% were women and mean baseline haemoglobin was 8.6 g/dL. Type 2 infarction accounted for 55.8%, type 1 for 41.7%, and 81.3% had non-ST-segment elevation infarction. Almost half were in an intensive or coronary care unit at randomisation; 13.7% had received mechanical ventilation. Before allocation, 35.3% had received red cells and 28.6% had undergone revascularisation. The mean interval from infarction to randomisation was 3.2 days.^{1,4}

The strategies produced substantial separation. Mean haemoglobin on day 1 was 8.8 versus 10.1 g/dL and on day 3 was 8.9 versus 10.5 g/dL, restrictive versus liberal. Mean red-cell use was 0.7 ± 1.6 versus 2.5 ± 2.3 units, with 1,237 versus 4,325 units administered overall. No post-randomisation transfusion was given to 66.3% versus 5.1%. Protocol discontinuation occurred in 46 patients (2.6%) versus 241 (13.7%). Liberal-group discontinuations included clinical concerns, patient or provider preference, and logistical problems. Median hospital stay after randomisation was five days in both groups.¹

The primary outcome occurred in 295/1,749 restrictive patients (16.9%) and 255/1,755 liberal patients (14.5%). The prespecified site-adjusted, multiply imputed risk ratio was 1.15 (95% CI 0.99–1.34; $P=0.07$). The crude risk ratio was 1.16 (95% CI 1.00–1.35); a post hoc analysis adjusting for baseline prognostic variables gave 1.16 (95% CI 1.00–1.36). Neither replaces the prespecified primary estimate. Death occurred in 173 (9.9%) versus 146 (8.3%; RR 1.19, 95% CI 0.96–1.47), and recurrent infarction in 149 (8.5%) versus 126 (7.2%; RR 1.19, 95% CI 0.94–1.49). The expanded secondary composite occurred in 19.6% versus 17.4% (RR 1.13, 95% CI 0.98–1.29).¹

Cardiac death occurred in 97 restrictive patients (5.5%) versus 56 liberal patients (3.2%; RR 1.74, 95% CI 1.26–2.40). Heart failure occurred in 5.8% versus 6.3% (RR 0.92, 95% CI 0.71–1.20). Transfusion-associated cardiac overload, the term used for TACO in the report, occurred in eight (0.5%) versus 23 (1.3%; RR 0.35, 95% CI 0.16–0.78). Transfusion-related acute lung injury was reported in 0/1,748 versus 6/1,754 patients. Venous thromboembolism occurred in 1.5% versus 1.9% (RR 0.77, 95% CI 0.46–1.27).^{1,4}

In the prespecified type 1 infarction subgroup, primary events occurred in 18.2% versus 13.8% (RR 1.32, 95% CI 1.04–1.67). Corresponding rates in type 2 infarction were 15.8% versus 15.1% (RR 1.05, 95% CI 0.85–1.29). The authors described the overall findings as consistent across prespecified subgroups.¹

Critique

MINT did not establish the superiority of liberal transfusion, but it also did not establish that restriction is equally effective or acceptably safe. The primary interval extends from a negligible relative advantage for restriction to a clinically important disadvantage. This asymmetry matters: the result is not an estimate close to exact equality with substantial potential benefit in either direction. Nevertheless, a P value of 0.07 cannot be relabelled a positive primary result because it lies near 0.05. The appropriate interpretation preserves both the concern about restrictive treatment and the remaining uncertainty about liberal benefit.¹

The trial achieved its recruitment target, generated a clinically relevant treatment contrast and obtained nearly complete follow-up. Its result should not be dismissed simply as an unsuccessful attempt to recruit enough patients. The observed between-group difference was smaller than the effect anticipated in the sample-size calculation. A smaller benefit could still be worthwhile for a common treatment decision in a high-risk population, but the primary interval, rather than retrospective power calculations, describes what this trial can support. The mortality and reinfarction components both pointed towards liberal treatment; their intervals remained

compatible with no effect. Moreover, components and expanded composites share patients and events, so their agreement is not independent replication.^{1,2}

The analysis-plan history further prevents interpreting the result as an endorsement of restriction. The original design specified superiority and no futility analysis. During a funding extension, a 2022 amendment added futility assessment and a fallback non-inferiority comparison, using an upper risk-ratio margin of 1.15 for restriction. The monitoring committee subsequently recommended continuing to the original target. The final plan retained that additional hypothesis, although the primary publication presents superiority analyses. An upper confidence limit of 1.34 does not satisfy the added non-inferiority criterion. This observation does not convert the trial into an original non-inferiority design; it reinforces that failure to demonstrate liberal superiority cannot validate restriction.²

The transfusion contrast also needs accurate description. MINT was not a rigid comparison of haemoglobin targets of 7 and 10 g/dL. Restrictive patients could be transfused below 8 g/dL, could receive blood for refractory angina, and had mean haemoglobin close to 9 g/dL during the initial follow-up. Many never required transfusion. Consequently, the trial evaluates a clinical policy of deferral with rescue provisions against proactive correction, not a requirement to maintain every restrictive patient at 7 g/dL. Extending its results to a stricter policy without the ischaemic-symptom exception would go beyond the intervention tested.^{1,2}

There was considerably more protocol discontinuation in the liberal group, and achievement of its haemoglobin threshold at discharge was reported as 86.3%. That limits conclusions about perfect target adherence, but the difference in blood use and achieved haemoglobin argues against wholesale failure to deliver the intervention. Reasons for stopping included fluid overload and other clinical concerns: these are relevant consequences of attempting the strategy, not merely inconvenient departures from it. Excluding such patients in a simple per-protocol comparison could select those best able to tolerate transfusion. The trial does not establish that stricter adherence would necessarily increase benefit without increasing harm.¹

Its population was broader than the phrase acute coronary occlusion might suggest. More than half had type 2 myocardial infarction, where oxygen supply–demand imbalance rather than acute atherothrombotic plaque disruption defines the mechanism. This makes the study relevant to complex medical and critical care patients, while raising the possibility that different causes of ischaemia respond differently. It does not demonstrate that type 2 patients diluted an otherwise proven benefit. The type 1 subgroup interval excluded one, whereas the type 2 interval did not; that distinction is

not itself a test of interaction. The supplied outcome displays provide no interaction estimate establishing a different treatment effect. Subgroup findings therefore do not validate a type-specific transfusion rule.^{1,2}

Timing and preceding treatment further define the question. Randomisation occurred a mean of 3.2 days after the infarction, more than one-third had already received red cells, and some had undergone revascularisation. These exposures preceded allocation and do not invalidate the randomised comparison. They do mean that MINT concerns subsequent inpatient transfusion management among patients still meeting its criteria, rather than the first transfusion decision at presentation or a strategy of never transfusing. It also cannot determine whether initiating the same policy earlier would produce a different result.^{2,4}

The primary outcome combined two serious events and had an important protection against ascertainment bias: recurrent infarction was assessed centrally without knowledge of allocation, using scheduled surveillance as well as clinically suspected events. Open treatment could still influence additional investigations, revascularisation or other care, and local classification of the qualifying infarction was not independently adjudicated. Those limitations should be distinguished from an unblinded adjudication of every primary event, which is not what occurred. The absolute mortality component is also less dependent on diagnostic interpretation than attribution of the cause of death.¹

The cardiac-death difference is striking but cannot carry a definitive survival claim. Although cause of death was collected prospectively, cardiac death was not designated within the primary, secondary or tertiary outcome hierarchy, was not centrally adjudicated, and was one of many comparisons without multiplicity adjustment. Knowledge of transfusion allocation could influence attribution; 40 deaths had an undetermined cause. The all-cause mortality estimate is consequently the more secure basis for assessing survival. The cardiac signal supports concern and further investigation, but cannot replace that less conclusive result.^{1,4}

Safety requires similar discrimination. Overall heart failure did not clearly differ, whereas attributed TACO was more frequent with liberal transfusion and several other transfusion reactions occurred predominantly in that group. Open-label attribution may contribute to differences in reported complications, but greater exposure is also a credible source of genuine harm. Similar heart-failure rates do not erase the TACO finding, and infrequent events cannot establish freedom from serious transfusion injury. Importantly, the liberal strategy incorporated diuresis, dialysis timing and clinical discretion; its safety should not be extrapolated to indiscriminate transfusion regardless of fluid tolerance.^{1,4}

The supplied analysis separates crude event counts from the imputed primary result. For crude risks, unobserved follow-up was treated as having no further event, rather than restricting analysis to participants with complete data. The primary model instead incorporated imputation for 57 participants, with more missing follow-up in the liberal group. Agreement between estimates is reassuring, but the rounded lower confidence limit of 1.00 in the crude analysis is not evidence that an alternative analysis conclusively succeeded. The prespecified model remains the principal result; imputation still relies on assumptions about unobserved outcomes.^{1,4}

Finally, physician agreement that either strategy was appropriate selected a population with clinical equipoise. The flow diagram begins with patients who consented, not every patient screened, limiting assessment of how many eligible patients were not approached or enrolled. Although multinational, recruitment was concentrated in the United States and Canada. Uncontrolled haemorrhage and planned cardiac surgery were excluded, and the primary assessment was at 30 days. The trial therefore offers substantial but bounded evidence for acute myocardial infarction with anaemia. It does not identify an optimal intermediate threshold, establish a long-term survival advantage, or justify ignoring the extra blood use and fluid burden when considering a liberal strategy.^{1,2,4}

Body of Evidence

The preceding evidence is best separated into trials of transfusion in patients with cardiovascular risk and trials conducted during acute coronary disease. A patient with stable coronary disease recovering from surgery does not necessarily have the same balance between oxygen supply, myocardial demand and transfusion risk as a patient with a recent infarction. Similar numerical transfusion thresholds across studies are therefore insufficient to make their results interchangeable.^{1,3}

FOCUS enrolled 2,016 patients aged at least 50 years after hip-fracture surgery, with cardiovascular disease or risk factors and haemoglobin below 10 g/dL. Liberal treatment maintained haemoglobin at least 10 g/dL; restriction permitted transfusion for symptoms or below 8 g/dL. Death or inability to walk independently at 60 days occurred in 34.7% with restriction and 35.2% with liberal treatment; the reported liberal-versus-restrictive odds ratio was 1.01 (95% CI 0.84–1.22). Median transfusion use was zero versus two units. This supported avoiding routine liberal transfusion in that postoperative population. However, recognised myocardial infarction within the previous 30 days and symptomatic anaemia were exclusions, preventing its direct use as evidence for the MINT population.³

Among the early acute-infarction studies, CRIT was very small. The MINT protocol describes 45 patients assigned to haematocrit triggers of 24% or 30%, with the in-hospital composite of death, recurrent infarction or new or worsening heart failure occurring in three restrictive and eight liberal patients. This account is important because the early randomised literature did not uniformly favour liberal treatment. However, an endpoint including heart failure, very small event numbers and different transfusion triggers prevent treating CRIT as a precise contradiction of the later MINT result. It principally illustrates how uncertain the starting evidence was.²

The 2013 MINT pilot randomised 110 patients with acute coronary syndromes or stable coronary disease undergoing catheterisation. Liberal treatment maintained haemoglobin at least 10 g/dL; restriction permitted transfusion below 8 g/dL or for symptoms. The 30-day composite of death, infarction or unscheduled revascularisation occurred in 14 restrictive and six liberal patients (RR 2.38, 95% CI 0.99–5.73; reported $P=0.054$). Seven versus one patient died. However, restrictive patients were approximately seven years older on average, and age adjustment weakened the primary signal. Recruitment stopped below its intended sample for feasibility reasons. The pilot supported a larger trial, not an established mortality benefit; its mixed coronary population and wider composite also differed from the definitive study.⁵

REALITY subsequently randomised 668 patients with acute myocardial infarction and haemoglobin of 7–10 g/dL in France and Spain. Its restrictive trigger was 8 g/dL or lower; the liberal trigger was 10 g/dL or lower, with a post-transfusion target of at least 11 g/dL. The 30-day primary clinical composite comprised death, stroke, recurrent infarction or emergency revascularisation. Unlike MINT's original design, REALITY tested non-inferiority, allowing a relative-risk margin of 1.25. The evidence available when MINT was published supported the prespecified 30-day non-inferiority conclusion. However, that does not mean the two strategies were proven equivalent, nor that a smaller excess risk from restriction would be unimportant.^{1,6}

The MINT investigators also cited REALITY's one-year follow-up: event rates were similar, but restriction did not meet the prespecified non-inferiority threshold at that later time point. This illustrates why a conclusion depends on its assessment window, margin and precision, not simply which group has fewer observed events. MINT's substantially larger sample and narrower primary composite added considerable information, but its result should be interpreted alongside these differences rather than portrayed as a straightforward reversal of an earlier definitive answer.¹

Taken together, the evidence through 2023 leaves acute myocardial infarction distinct from many other transfusion settings. MINT provides a substantial signal that routine

restriction may sacrifice clinically important benefit, while falling short of establishing liberal superiority. Its findings make blanket reassurance about a 7–8 g/dL strategy difficult to sustain, but do not establish an exact optimal haemoglobin target or eliminate transfusion-related harm. The resulting clinical judgement is a reasoned preference under uncertainty, with attention to active ischaemia and fluid tolerance, rather than a claim that the primary hypothesis was conclusively confirmed.¹

Should we use liberal transfusion in anaemic patients with acute myocardial infarction?

Consider it. Superiority was not established, but clinically important harm from restriction remains possible.

References

1. Carson JL, Brooks MM, Hébert PC, Goodman SG, Bertolet M, Glynn SA, et al. Restrictive or Liberal Transfusion Strategy in Myocardial Infarction and Anemia. *N Engl J Med* 2023;389(26):2446–2456.
2. Carson JL, Brooks MM, Hébert PC, et al. Protocol and statistical analysis plans for: Restrictive or Liberal Transfusion Strategy in Myocardial Infarction and Anemia. *N Engl J Med* 2023;389(26):2446–2456.
3. Carson JL, Terrin ML, Noveck H, Sanders DW, Chaitman BR, Rhoads GG, et al. Liberal or Restrictive Transfusion in High-Risk Patients after Hip Surgery. *N Engl J Med* 2011;365(26):2453–2462.
4. Carson JL, Brooks MM, Hébert PC, et al. Supplementary appendix to: Restrictive or Liberal Transfusion Strategy in Myocardial Infarction and Anemia. *N Engl J Med* 2023;389(26):2446–2456.
5. Carson JL, Brooks MM, Abbott JD, Chaitman B, Kelsey SF, Triulzi DJ, et al. Liberal versus restrictive transfusion thresholds for patients with symptomatic coronary artery disease. *Am Heart J* 2013;165(6):964–971.e1.
6. Ducrocq G, Gonzalez-Juanatey JR, Puymirat E, Lemesle G, Cachanado M, Durand-Zaleski I, et al. Effect of a Restrictive vs Liberal Blood Transfusion Strategy on Major Cardiovascular Events Among Patients With Acute Myocardial Infarction and Anemia: The REALITY Randomized Clinical Trial. *JAMA* 2021;325(6):552–560.

ECLS-SHOCK

Thiele H, Zeymer U, Akin I, Behnes M, Rassaf T, Mahabadi AA, et al. Extracorporeal Life Support in Infarct-Related Cardiogenic Shock. N Engl J Med 2023;389(14):1286–1297.

Introduction

Restoring coronary blood flow does not necessarily restore an effective circulation. Patients with acute myocardial infarction complicated by cardiogenic shock may remain profoundly hypotensive and hypoperfused despite revascularisation, catecholamines and intensive care. Contemporary mortality of approximately 40–50% at 30 days leaves substantial scope for improvement, but also creates pressure to adopt treatments before their effect on patient-centred outcomes has been established.¹

Venoarterial extracorporeal membrane oxygenation, also termed extracorporeal life support (ECLS), offers an intuitively compelling response. Venous blood is drained, oxygenated and returned to the arterial circulation, providing both circulatory and respiratory support while the heart recovers. Smaller systems and percutaneous cannulation have facilitated its use in the catheterisation laboratory. Unlike an intervention directed solely at the coronary lesion, ECLS can sustain systemic perfusion during severe pump failure. These capabilities explain its attraction; they do not establish that routinely initiating it improves survival.^{1,2}

The potential disadvantages are equally relevant. Large arterial cannulae can compromise limb perfusion, and vascular instrumentation and anticoagulation expose patients to bleeding. Retrograde arterial flow may increase left ventricular afterload, potentially impeding ejection and requiring additional unloading. Furthermore, restoring circulation cannot necessarily reverse established hypoxic brain injury or advanced organ damage. A physiological benefit may therefore coexist with procedural harm or arrive too late to alter the outcome.^{1,3}

The clinical question is consequently more precise than whether ECLS can support a failing circulation. Should it be initiated routinely and early in infarct-related cardiogenic shock, or should initial treatment consist of revascularisation and medical care, with mechanical support reserved for deterioration? ECLS-SHOCK tested the former strategy in patients with substantial hypoperfusion undergoing planned early revascularisation. Its hypothesis was that adding early ECLS would reduce death from any cause at 30 days.¹

Synopsis

ECLS-SHOCK was an investigator-initiated, multicentre, open-label, randomised superiority trial conducted at 44 centres in Germany and Slovenia between June 2019 and November 2022. Funding came from the Else Kröner Fresenius Foundation, German Heart Research Foundation and Helios Health Institute. An independent data and safety monitoring board reviewed the trial, and a clinical-events committee assessed outcomes. A stepwise consent process accommodated patients unable to consent during shock, with subsequent patient or representative agreement.^{1,3}

Eligible patients were aged 18–80 years, had acute myocardial infarction and planned early percutaneous coronary intervention (PCI) or coronary artery bypass grafting. Shock required systolic pressure below 90 mmHg for more than 30 minutes, or catecholamines to maintain systolic pressure above 90 mmHg, together with arterial lactate above 3 mmol/L and impaired perfusion. The latter required altered mental status, cold or clammy extremities, or urine output below 30 mL/hour. Exclusions included shock onset more than 12 hours earlier, resuscitation exceeding 45 minutes, mechanical complications causing shock, non-infarct causes of shock, peripheral arterial disease preventing cannulation, pregnancy and another illness limiting survival to less than six months. The original upper age limit of 75 years increased to 80 during recruitment.^{1,2}

After diagnostic coronary angiography, allocation was performed centrally through a web-based system, using variable blocks, stratification by centre and a 1:1 ratio. Both groups received early revascularisation, preferably PCI, and medical and intensive care treatment. The protocol favoured culprit-lesion PCI rather than routine immediate multivessel intervention. Treating clinicians were not blinded.^{1,2}

Intervention patients were to receive percutaneous femoral ECLS during the index catheterisation. Cannulation before revascularisation was preferred provided it did not delay reperfusion by more than 20 minutes. An antegrade distal perfusion sheath was strongly recommended. Intensive care recommendations included a mean arterial pressure target of 65 mmHg, avoidance of excessive afterload, right-arm arterial monitoring, regular limb assessment and unfractionated heparin targeting an activated partial thromboplastin time of 60–80 seconds or activated clotting time of 250 seconds. Initial flow was to be maximised within haemodynamic, cannula and inlet-pressure constraints.^{2,3}

Left ventricular unloading was selective, not routine. Prespecified triggers included absent arterial pulsatility or aortic valve opening, a left ventricular outflow tract velocity-time measurement below 10 cm and progressive ventricular enlargement. Options included an intra-aortic balloon pump, microaxial pump or catheter venting.

Weaning required stable pressure, native ejection with minimal inotropic support, no important hypoxaemia, lactate below 2 mmol/L and central venous oxygen saturation above 65%. Recommended flow reductions were 0.5–1 L/min over four to six hours, aiming for 1–1.5 L/min before removal.³

Control patients received medical treatment without routine ECLS. Crossover was discouraged, but escalation to another support device was permitted for impending haemodynamic collapse, a lactate rise exceeding 3 mmol/L within six hours, or a 50% increase in vasopressor requirement to maintain mean pressure above 65 mmHg. Thus, the comparator incorporated defined opportunities for additional mechanical support rather than prohibiting it under all circumstances.^{1,2}

The primary outcome was all-cause mortality at 30 days. Secondary outcomes included haemodynamic stabilisation, durations of catecholamine treatment, ventilation and intensive care, renal replacement therapy, reinfarction, repeat revascularisation and heart-failure readmission. Serial lactate, renal function, troponin and illness-severity measurements were collected. Safety outcomes comprised Bleeding Academic Research Consortium (BARC) type 3–5 bleeding, peripheral ischaemic vascular complications requiring intervention or surgery, and stroke or systemic embolisation. A poor neurological outcome defined by Cerebral Performance Category (CPC) 3–4 at 30 days was a post hoc endpoint.^{1,3}

The sample-size calculation assumed mortality of 35% with ECLS and 49% with control treatment: a 14-percentage-point absolute reduction. A sample of 394 provided 80% power, with a final two-sided significance threshold of 0.048 after allowing for an interim analysis. Recruitment of 420 allowed for 6% attrition. The interim review did not recommend stopping. The primary comparison followed allocation, using a chi-square test and relative risk with 95% confidence intervals. Per-protocol and as-treated analyses assessed consistency. Secondary and subgroup intervals were not adjusted for multiple comparisons.^{1,2,3}

Of 877 patients screened, 420 were randomised. Three subsequently declined consent, leaving 209 ECLS and 208 control patients, with complete primary-outcome follow-up. Median age was 63 years, 81.3% were men, approximately two-thirds had ST-segment elevation infarction, and median left ventricular ejection fraction was 30%. Median lactate was 6.9 mmol/L and pH approximately 7.2. Previous cardiopulmonary resuscitation occurred in 77.7%, with a median longest resuscitation episode of 20 minutes. Nearly all revascularisation was by PCI.^{1,3}

ECLS was delivered to 192/209 intervention patients (91.9%). Of these, 42 (21.9%) started before revascularisation, 50 (26.0%) during it and 100 (52.1%) afterwards. Seventeen did not receive ECLS, including four who died before insertion. Median support duration was 2.7 days; distal perfusion sheaths were used in 183/192 patients (95.3%), and active unloading in 11/191 (5.8%) according to the main treatment table. Control-group ECLS occurred in 26/208 patients (12.5%): 22 within 24 hours and four later. Among the 182 control patients without ECLS, 28 received another support device, comprising 27 microaxial pumps and one balloon pump.^{1,3}

Thirty-day mortality was 100/209 (47.8%) with ECLS and 102/208 (49.0%) with control treatment: relative risk 0.98 (95% CI 0.80–1.19; P=0.81). Per-protocol and as-treated estimates were similar. Prespecified subgroup analyses considered age, sex, diabetes, infarction characteristics and admission lactate; analysis by previous resuscitation was post hoc. None identified a convincing differential benefit.^{1,3}

Renal replacement therapy occurred in 17 patients (8.1%) versus 29 (13.9%; RR 0.58, 95% CI 0.33–1.03), and repeat revascularisation in 18 (8.6%) versus 22 (10.6%; RR 0.81, 95% CI 0.45–1.47). Two patients in each group had reinfarction. Median time to haemodynamic stabilisation was 3.1 days in both groups. Median ventilation duration was seven versus five days; the reported Hodges–Lehmann difference estimate was one day (95% CI 0–2). Median intensive care duration was ten versus eight days. The reported survivor-based CPC 3–4 counts were 27/109 (24.8%) versus 24/106 (22.6%).^{1,3}

Moderate or severe bleeding occurred in 49 ECLS patients (23.4%) versus 20 controls (9.6%; RR 2.44, 95% CI 1.50–3.95). Peripheral vascular complications requiring intervention occurred in 23 (11.0%) versus eight (3.8%; RR 2.86, 95% CI 1.31–6.25). Stroke or systemic embolisation occurred in eight (3.8%) versus six (2.9%; RR 1.33, 95% CI 0.47–3.76). Refractory shock accounted for approximately half the deaths in each group and brain injury for approximately one-quarter.^{1,3}

Critique

The most important distinction is between routine early support and selective rescue. ECLS-SHOCK does not compare unrestricted access to ECLS with a system in which it is unavailable. Its control strategy allowed escalation to other devices, and some patients received ECLS after deterioration. The result therefore challenges routinely adding ECLS around revascularisation, not every conceivable indication for rescue support. Equally, the availability of rescue cannot be used to claim that rescue itself improved survival: those decisions were not randomised.^{1,2}

Central allocation, multicentre recruitment, a patient-centred primary endpoint and complete ascertainment among the 417 retained participants strengthen the main comparison. The three consent-related exclusions mean that not every randomised participant was analysed, but attrition was small. Mortality is less vulnerable to diagnostic subjectivity than a composite including the clinician's decision to implant another device. These features make the neutral primary result difficult to dismiss as merely a problem of outcome measurement. The final analysis plan was dated January 2023, after recruitment ended, but retained the primary mortality endpoint and chi-square comparison while changing several effect measures from odds ratios to risk ratios. This chronology should not be confused with changing the primary hypothesis.^{1,2,3}

Precision still matters. The confidence interval is compatible with approximately a 20% relative mortality reduction or a 19% increase; this is not an equivalence trial. However, the assumed 14-point absolute benefit was ambitious. It arose from a meta-analysis of only 95 patients in two retrospective comparisons. Selecting the lower confidence limit of a large observational benefit does not remove confounding or ensure a realistic randomised effect. The observed control mortality closely matched the planning assumption, but the anticipated large treatment benefit did not emerge. Smaller clinically important effects remain possible.^{1,2,3}

Non-delivery and crossover could weaken the contrast, although ECLS exposure remained markedly different between groups. The apparently conflicting crossover totals also require attention to timing. The appendix counts crossover within 24 hours: 17 intervention patients without ECLS and 22 controls receiving it. Four additional controls received ECLS later. The per-protocol and as-treated displays use that early window, rather than simply regrouping every patient by any subsequent exposure. More importantly, excluding or reallocating patients according to deterioration, recovery or failure to survive until cannulation breaks randomisation. Their similar results are reassuring sensitivity findings, not unbiased estimates of the treatment's biological efficacy.^{1,2,3}

Timing is another boundary. Only about one-fifth of supported intervention patients underwent cannulation before revascularisation, despite the protocol's preference. The trial therefore does not test mandatory pre-PCI ECLS in every eligible patient. A benefit from a different timing strategy remains a hypothesis, not an explanation established by these data. Delaying coronary reperfusion to ensure device insertion also carries a competing concern, explicitly recognised by the protocol's 20-minute limit. Selecting patients retrospectively by implantation timing would not resolve that trade-off reliably.^{1,2}

The high prevalence of preceding cardiac arrest raises a plausible competing mechanism of death: systemic perfusion cannot reverse all established brain injury. Nevertheless, these were patients with infarction and shock after resuscitation, not a randomised trial of ECLS initiated during refractory arrest. Brain injury accounted for approximately one-quarter of deaths, whereas refractory shock remained the leading cause. The post hoc resuscitation subgroup did not identify benefit in patients without prior arrest, and that subgroup was small. It is therefore inappropriate either to dismiss the trial as dominated by irreversible neurological injury or to declare a validated non-arrest population that benefits.^{1,3}

The shock-stage descriptions need particular care. Although presented in the baseline table, the modified Society for Cardiovascular Angiography and Interventions classification was assigned post hoc. Stages C and D incorporated the direction of lactate change at eight hours, and stage D could incorporate death during those hours. These are not purely pretreatment characteristics and may themselves be influenced by allocation. The apparent imbalance in stage D cannot consequently be interpreted as an ordinary baseline imbalance. Nor would a treatment comparison conditioned on these reconstructed stages establish which presenting shock phenotype should receive ECLS.¹

Only 5.8% underwent active left ventricular unloading. This leaves open whether a routinely combined support-and-unloading strategy would behave differently, especially given ECLS-related afterload. However, the protocol specified selective triggers, and the supplied data do not establish how many unsupported ventricles met those triggers. A low unloading rate alone is therefore insufficient evidence of protocol failure. Additional vascular access and devices also introduce harm. A different combination requires its own randomised evaluation rather than being presumed to rescue the negative finding.^{1,3}

The observed complications are integral to the treatment effect, not incidental technical details. Substantially more bleeding and limb complications occurred despite frequent distal perfusion and relatively small arterial cannulae. The secondary intervals were unadjusted for multiplicity, but the magnitude, direction and physiological plausibility of these prespecified safety findings are concerning. It remains unproven that they exactly counterbalanced a survival benefit. Conversely, the absence of a clear stroke difference does not establish neurological safety, and the uncertain renal-replacement estimate cannot be promoted to a confirmed organ-protective effect.^{1,3}

Open treatment could influence co-interventions and decisions about ventilation or treatment limitation. Dobutamine was more frequently used with ECLS, while recorded temperature management was less frequent. The authors suggested that temperature

control through the circuit might have been under-recorded, but this was not demonstrated. Duration endpoints also require caution: early death can shorten ventilation or intensive care. The investigators did not incorporate death as a competing event in those comparisons because mortality frequency and timing were similar. That rationale does not make a shorter duration synonymous with better recovery.^{1,3}

Neurological interpretation is further restricted by survivor-based assessment and the post hoc CPC 3–4 definition. A comparison among survivors is not the same as survival with an acceptable neurological outcome in all randomised patients. The protocol amendments and final outcome documentation also differ concerning day-30 CPC assessment. These limitations matter to claims about neurological recovery, but do not invalidate the separately ascertained all-cause mortality endpoint.^{1,2,3}

Finally, applicability is concentrated in selected infarct-related shock treated in predominantly German catheterisation centres. Women were under-represented, patients over 80 were excluded, and prolonged resuscitation, mechanical complications and other causes of shock were outside the trial. Thirty-day follow-up does not describe later function or quality of life. Within its actual population, however, the balance of an unproven survival advantage and increased complications argues against routine early ECLS. Uncertainty about a narrower indication is a reason for further evidence, not for assuming that the overall result can be ignored.^{1,3}

Body of Evidence

The preceding literature separates the value of treating the infarction from the value of adding a circulatory device. In its account of the SHOCK trial, the ECLS-SHOCK protocol describes a comparison of early revascularisation with initial medical stabilisation. Thirty-day mortality was 46.7% versus 56.0% ($P=0.11$), with a mortality advantage emerging at six months. That experience cautions against equating every neutral short-term result with lifelong ineffectiveness. It does not, however, justify assuming that an unproven ECLS benefit will appear later; the intervention and its harms are different.²

The same protocol summarises IABP-SHOCK II, which randomised 600 patients with infarction-related shock undergoing early revascularisation to an intra-aortic balloon pump or medical treatment alone. No mortality benefit was demonstrated at 30 days or 12 months. This weakened the case for routine balloon counterpulsation and increased interest in more active support. However, failure of one device does not establish efficacy of a more powerful alternative. The amount of flow delivered, its effect on

ventricular loading and the associated complications must all be considered, rather than ranking devices solely by haemodynamic capability.^{1,2}

The observational evidence underlying the ECLS-SHOCK sample-size assumption was much more optimistic. Its two small retrospective comparisons suggested a substantial survival advantage for ECLS over balloon support. These comparisons could not ensure that patients receiving each treatment were otherwise comparable. Treatment selection, cannulation feasibility and the opportunity to survive until implantation can all complicate interpretation. The randomised trial therefore tested a proposition suggested by these data; it did not simply reproduce an already established treatment effect with a different statistical method.^{2,3}

Thiele and colleagues describe three earlier randomised ECLS studies. The first, the small trial by Brunner and colleagues in infarct-related cardiogenic shock, did not demonstrate improvement in left ventricular ejection fraction at 30 days. Such a physiological outcome addresses a different question from mortality and may be difficult to interpret when many patients do not survive to assessment. Its limited size could not resolve the survival question that ECLS-SHOCK was designed to address.¹

ECMO-CS enrolled 122 patients at four centres with severe or deteriorating cardiogenic shock, including both infarction-related shock and decompensated heart failure. As described in the ECLS-SHOCK paper and accompanying editorial, it compared immediate ECLS with initial conservative treatment permitting later ECLS when required. There was no demonstrated improvement in its 30-day composite of death, resuscitated cardiac arrest or another mechanical support device. This supports a distinction between routine and selective support, but the mixed aetiologies and inclusion of a treatment-escalation decision within the endpoint prevent direct interchange with ECLS-SHOCK's infarct-specific mortality comparison.^{1,4}

EURO SHOCK was more directly concerned with myocardial infarction complicated by cardiogenic shock. The ECLS-SHOCK investigators describe premature termination after only 35 of the planned 428 patients had been randomised. This severely limited the conclusions that could be drawn about mortality. Its importance is therefore not as a definitive positive or negative counterweight, but as another illustration of the difficulty of assembling sufficiently precise randomised evidence for this intervention.¹

A contemporaneous individual-patient-data meta-analysis combined four randomised trials involving 567 patients with infarct-related cardiogenic shock. Early VA-ECMO did not demonstrate lower 30-day mortality (odds ratio 0.93, 95% CI 0.66–1.29). Major bleeding (OR 2.44, 95% CI 1.55–3.84) and peripheral ischaemic complications (OR 3.53,

95% CI 1.70–7.34) were more frequent. Prespecified subgroup analyses did not establish a population with benefit. These are odds ratios, not ECLS-SHOCK's risk ratios. Importantly, the synthesis includes ECLS-SHOCK and is not independent replication of its results.⁵

Leopold and Taichman interpreted the trial as an argument against routine early ECLS, while acknowledging that the study did not identify which individual patients might benefit. That is a proportionate distinction between rejecting indiscriminate use and declaring the technology universally ineffective. Through the end of 2023, the randomised evidence does not support routine early ECLS in infarct-related shock. The remaining questions concern selecting patients, timing support and reducing its complications, without presuming that a preferred untested strategy has already been validated.^{1,4,5}

Should we use routine early ECLS in infarct-related cardiogenic shock?

Not routinely. No survival benefit was demonstrated, and bleeding and vascular complications increased.

References

1. Thiele H, Zeymer U, Akin I, Behnes M, Rassaf T, Mahabadi AA, et al. Extracorporeal Life Support in Infarct-Related Cardiogenic Shock. *N Engl J Med* 2023;389(14):1286–1297.
2. Thiele H, Zeymer U, Akin I, et al. Protocols, amendments and statistical analysis plans for: Extracorporeal Life Support in Infarct-Related Cardiogenic Shock. *N Engl J Med* 2023;389(14):1286–1297.
3. Thiele H, Zeymer U, Akin I, et al. Supplementary appendix to: Extracorporeal Life Support in Infarct-Related Cardiogenic Shock. *N Engl J Med* 2023;389(14):1286–1297.
4. Leopold JA, Taichman DB. Routine Early ECLS in Infarct-Related Cardiogenic Shock? *N Engl J Med* 2023;epublished 26 August. doi:10.1056/NEJMe2309395.
5. Zeymer U, Freund A, Hochadel M, Ostadal P, Belohlavek J, Rokyta R, et al. Venoarterial extracorporeal membrane oxygenation in patients with infarct-related cardiogenic shock: an individual patient data meta-analysis of randomised trials. *Lancet* 2023;402(10410):1338–1346.



Resuscitation Trials

TAME

Eastwood G, Nichol AD, Hodgson C, Parke RL, McGuinness S, Nielsen N, et al. Mild Hypercapnia or Normocapnia after Out-of-Hospital Cardiac Arrest. *N Engl J Med* 2023;389(1):45–57.

Introduction

Restoration of circulation after cardiac arrest does not necessarily restore adequate cerebral perfusion. Hypoxic–ischaemic brain injury remains a major determinant of subsequent death and disability, and persistent hypoperfusion offers a potential therapeutic target after the initial resuscitation has succeeded. An intervention that increases cerebral blood flow without requiring a new medicine or invasive device would therefore be particularly attractive.¹

Arterial carbon dioxide tension is an important regulator of cerebrovascular tone. Increasing PaCO₂ causes cerebral vasodilatation, and the physiological response appears to remain present after cardiac arrest. This creates a plausible argument for deliberately maintaining mild hypercapnia rather than merely avoiding hypocapnia. However, greater cerebral blood flow is not synonymous with neurological recovery. The intervention must act early enough, improve oxygen delivery to tissue capable of recovery, and avoid offsetting adverse effects.^{1,2}

The evidence preceding TAME was encouraging but incomplete. Observational studies associated hypercapnia with better outcomes, a small physiological study demonstrated improved cerebral oxygenation, and a phase II randomised trial suggested attenuation of neuronal injury. None established that deliberately targeting mild hypercapnia would improve survival with an acceptable neurological outcome. The contemporary normocapnia recommendation consequently rested alongside a biologically plausible alternative whose clinical value remained uncertain.^{1,2}

The Targeted Therapeutic Mild Hypercapnia after Resuscitated Cardiac Arrest trial tested whether this physiological rationale could translate into patient benefit. Its question was specific: would targeting PaCO₂ of 50–55 mmHg for 24 hours improve six-month neurological recovery compared with targeting 35–45 mmHg in adults remaining unconscious after out-of-hospital cardiac arrest? It was not a comparison of unrestricted permissive hypercapnia with aggressive hyperventilation, nor a trial of individually titrated cerebral perfusion treatment.¹

Synopsis

TAME was an investigator-initiated, international, multicentre, open-label, parallel-group superiority trial conducted in 63 ICUs across 17 countries between March 2018 and September 2021. Funding included Australian, Irish and New Zealand public research agencies. Consent was obtained from a surrogate or deferred according to local requirements, with continued participation discussed with patients who regained capacity. An independent committee monitored safety and reviewed the prespecified interim analysis.^{1,3}

Eligible patients were adults with a presumed cardiac or unknown cause of out-of-hospital cardiac arrest, sustained return of spontaneous circulation for at least 20 minutes without chest compressions, and persistent unconsciousness, defined by a Full Outline of Unresponsiveness motor score below 4. Screening had to occur within 180 minutes of return of spontaneous circulation. Important exclusions were an unwitnessed arrest with initial asystole, limitations on intensive care, admission temperature below 30°C, extracorporeal membrane oxygenation before return of spontaneous circulation, suspected pregnancy, intracranial bleeding, and severe chronic obstructive pulmonary disease requiring long-term home oxygen.³

Allocation was performed centrally through a web-based system using randomly permuted blocks and a 1:1 ratio, stratified by site and, whenever possible, concomitant enrolment in TTM2. Treating clinicians knew the assigned target. Neurological prognosticators and six-month outcome assessors were masked, as were the statisticians and authors during analysis. Alternative interpretations were written before the allocation code was revealed.¹

The intervention was a PaCO₂ target of 50–55 mmHg for 24 hours from randomisation. Respiratory rate adjustments to minute ventilation were guided by arterial blood gases and end-tidal carbon dioxide. The comparator used the same monitoring approach but targeted 35–45 mmHg. Both protocols specified a Richmond Agitation–Sedation Scale target of –4 and blood gases at baseline and every four hours, with additional measurements following an end-tidal carbon dioxide change exceeding 5 mmHg. Blood gases were assessed at 37°C using an alpha-stat approach, rather than corrected to the patient's actual temperature.³

Sedative choice, ventilator settings and neuromuscular blockade remained subject to clinical judgement. Short-term paralysis was recommended when required to achieve the hypercapnia target. Initial severe metabolic acidosis, defined by pH below 7.1 together with base excess below –6 mmol/L, delayed hypercapnia until corrected; this delay still counted within the 24-hour intervention. A gradual return to normocapnia over the

following four hours was recommended. Other treatment followed local practice, including temperature management; 370 patients were co-enrolled in TTM2.^{1,3}

For patients remaining in the ICU, the protocol specified blinded neurological prognostication at 96 hours or later. Confounding sedation and metabolic abnormalities were to be excluded, with clinical examination and electroencephalography supplemented by other investigations as appropriate. A poor neurological prognosis alone was insufficient grounds for withdrawing treatment before 96 hours, although earlier withdrawal was permitted when continued treatment was considered unethical for other reasons. The treating team retained responsibility for withdrawal decisions.³

The primary outcome was a favourable Glasgow Outcome Scale–Extended score of 5–8 at six months. The scale ranges from death at 1 to upper good recovery at 8; scores 5–6 represent moderate disability and 3–4 severe disability. A separate dichotomised outcome incorporated additional patients whose records permitted classification as favourable or unfavourable without a complete structured score. Secondary outcomes included death within six months, poor functional outcome defined by modified Rankin Scale scores 4–6, and patient-perceived health on the 0–100 EuroQol visual analogue scale. Deaths were assigned a visual analogue score of zero in the overall quality-of-life analysis.¹

The planned sample size of 1624 provided 90% power to detect an eight-percentage-point increase in favourable neurological outcome, from 50% to 58%, at a two-sided significance level of 0.05. Recruitment was increased to 1700 to allow for withdrawals and missing follow-up. An interim analysis after 850 patients used dichotomised modified Rankin outcomes and an O’Brien–Fleming stopping boundary of $P=0.005$. The trial continued, with a final primary significance threshold of 0.05.^{1,3}

The primary analysis followed assigned treatment after excluding patients who withdrew consent. Binary outcomes were analysed using mixed-effects binomial models with a log link, site as a random intercept and TTM2 co-enrolment as a fixed covariate. Missing outcomes were multiply imputed using baseline, prognostic and outcome information, assuming that missingness was explained by observed variables. Ten imputed datasets were combined. Sensitivity analyses considered ordinal GOS-E and patients not co-enrolled in TTM2. Secondary confidence intervals were not adjusted for multiplicity.^{1,3}

Of 5779 patients assessed, 1700 were randomised: 847 to mild hypercapnia and 853 to normocapnia. Twenty-four withdrew consent, leaving 1676 in the intention-to-treat population. Baseline data were available for 1668 because eight patients had consent

limited to allocation and mortality information. Mean age was approximately 61 years; 76.6% versus 81.2% were men. Approximately 88% had witnessed arrests, 81% received bystander cardiopulmonary resuscitation and 71% had an initial shockable rhythm. Median arrest-to-randomisation intervals were 154 versus 151 minutes. Initial mean PaCO₂ was already approximately 53 mmHg in both groups.^{1,3}

PaCO₂ separated during the first four hours and remained different thereafter. Time-weighted 24-hour averages were 49.8 versus 40.8 mmHg, with a median two hours to first attain the assigned range in each group. Among protocolised measurements, hypocapnia below 35 mmHg occurred in 139/4464 measurements (3.1%) versus 696/4397 (15.8%). The target was abandoned in 68/829 patients (8.2%) versus 25/839 (3.0%). Neuromuscular blockade was used in 67.7% versus 59.1%; median time from randomisation to extubation was 76.3 versus 70.3 hours.^{1,3}

Complete primary neurological assessments were available for 1548 patients, leaving 7.6% missing. Favourable outcome occurred in 332/764 patients (43.5%) with mild hypercapnia and 350/784 (44.6%) with normocapnia. The adjusted relative risk, incorporating multiple imputation, was 0.98 (95% CI 0.87–1.11; P=0.76). The unadjusted risk difference was –1.2 percentage points (95% CI –6.1 to 3.8). The expanded dichotomised assessment included 1594 patients and was consistent: 348/788 (44.2%) versus 365/806 (45.3%), adjusted relative risk 0.98 (95% CI 0.87–1.11).¹

Six-month mortality was 393/816 (48.2%) versus 382/832 (45.9%), relative risk 1.05 (95% CI 0.94–1.16). Poor modified Rankin outcome occurred in 407/762 (53.4%) versus 400/779 (51.3%), relative risk 1.05 (95% CI 0.95–1.15). Overall mean visual analogue scores were 35.8 versus 36.8, with an adjusted mean difference of 0.2 (95% CI –3.7 to 4.0). Sensitivity analyses were consistent with the primary result, without a convincing treatment difference across the prespecified subgroups or an identified interaction with TTM2 assignment.^{1,3}

Prespecified adverse events did not differ significantly. Pneumonia occurred in 40.8% versus 43.1%, sepsis in 12.1% versus 12.0%, and severe bleeding in 1.9% versus 1.7%. No raised intracranial pressure or seizure event necessitating a return to normocapnia was recorded, although seizures themselves occurred in both groups. One cerebral oedema event in the hypercapnia group was assessed as possibly intervention-related.^{1,3}

Critique

TAME provides a substantial test of a readily deliverable intervention with a plausible physiological basis. The important finding is not simply a P value above 0.05: there was

sustained carbon dioxide separation without corresponding improvement in neurological recovery, functional status or survival. Failure to deliver any meaningful exposure contrast is therefore an unconvincing explanation for the primary result. The trial directly challenges routine adoption of this particular 24-hour hypercapnia strategy.^{1,3}

Nevertheless, achieving different arterial values does not establish that cerebral perfusion or tissue oxygenation improved in every patient. The trial did not select patients by measured cerebral hypoperfusion or titrate treatment to an individual cerebral response. Different mechanisms and degrees of hypoxic–ischaemic injury could therefore coexist within the same target group. This limits mechanistic conclusions, but it is not a reason to dismiss the clinically relevant comparison: the proposed routine intervention was precisely to prescribe a carbon dioxide target without requiring specialised cerebral monitoring.^{1,2}

The distinction between average exposure and individual target attainment also matters. A time-weighted hypercapnia average of 49.8 mmHg lay just below the intended 50–55 mmHg range. The tightly grouped confidence intervals around mean values in the main figure should not be mistaken for uniformly tight control in individual patients; the supplementary figure displays much wider dispersion. Conversely, intermittent blood gases do not justify reconstructing continuous time in range. The reported difference is an effect of implementing these targets in practice, not of maintaining every patient continuously at exactly 52.5 or 40 mmHg.³

Timing is another boundary on interpretation. Randomisation occurred a median of about two and a half hours after arrest, and the first in-range measurement occurred a median two hours after randomisation. These medians cannot simply be added to create a measured combined interval, but they demonstrate that the intervention was not instituted immediately at reperfusion. The 24-hour clock also included the period needed to correct severe metabolic acidosis. Earlier treatment or a longer duration remains untested rather than an established explanation for failure.^{1,3}

The control was active normocapnia management, not an instruction to allow hypocapnia. Nevertheless, 15.8% of its protocolised measurements were below 35 mmHg, compared with 3.1% during targeted hypercapnia. The absence of benefit despite this additional separation is noteworthy, but cannot show that hypocapnia is harmless: patients were randomised to target strategies, not to individual hypocapnic episodes. Likewise, presenting these percentages as the proportions of patients exposed would misrepresent measurement-level data.¹

The primary outcome appropriately combined survival and longer-term neurological recovery, with blinded assessment reducing ascertainment bias. However, dichotomising GOS-E sacrifices distinctions within the favourable and unfavourable categories. A patient with moderate disability counted as favourable, not fully recovered. The ordinal sensitivity analyses were reassuringly consistent, although the three vegetative-state observations, all in the normocapnia group, required pooling with an adjacent category for model fitting. Neither the primary scale nor the visual analogue score substitutes for the detailed cognitive outcomes planned for separate publication.^{1,3}

The observed primary proportions and the reported relative risk require careful distinction. The former describe available structured assessments; the latter comes from an adjusted model with imputed missing outcomes. Additional record-based classification reduced missingness but was reported separately, not silently substituted for the original assessment. Agreement between these approaches strengthens the finding. Nevertheless, imputation relies on assumptions about unobserved outcomes after accounting for recorded information; it cannot guarantee removal of bias if follow-up depended on unmeasured disability or recovery.^{1,3}

The analysis-plan chronology also warrants transparency. Four amendments were made in 2022 after recruitment ended but, according to the supplied documents, before access to unblinded data. They included changing multiple imputation from a sensitivity analysis triggered by more than 5% missingness to its use whenever outcomes were missing. The clinical primary endpoint remained favourable six-month GOS-E, but the final publication's log-link model and unadjusted secondary confidence intervals differ from wording retained in the final analysis plan. These reporting differences merit clarification rather than an assumption that a different clinical endpoint was selected after viewing the result.^{1,2}

The confidence interval does not support the large benefit anticipated during planning. The unadjusted absolute interval extended from a 6.1-percentage-point reduction to a 3.8-point increase in favourable recovery, while the adjusted relative interval also remained compatible with modest benefit or harm. TAME was a superiority trial, not an equivalence or non-inferiority trial. Its results are therefore stronger grounds for withholding routine adoption than for declaring the two targets interchangeable under every clinical circumstance. Mortality was numerically higher with hypercapnia, but its confidence interval does not establish a causal increase in deaths.¹

Unblinded bedside care deserves consideration because the strategy influenced more than an arterial measurement. Neuromuscular blockade was more common and the target was abandoned more often with hypercapnia. These differences form part of the

practical treatment package being tested. Excluding patients who could not tolerate or sustain the target would risk selecting a prognostically different population. The May 2022 analysis-plan amendment explicitly acknowledged the absence of a prespecified adherence definition and deferred a per-protocol analysis; no such analysis should be assumed to rescue the result.^{2,3}

Delayed, blinded prognostication was an important protection against a self-fulfilling assessment of poor recovery. Actual withdrawal practice should nevertheless be distinguished from the intended pathway. Table S4 records poor neurological outcome among reasons for withdrawing treatment during days 1–3 in both groups, despite the restriction on neurological prognostication alone before 96 hours. Without patient-level circumstances these entries cannot be adjudicated as protocol violations, but the supplied data do not justify describing delayed withdrawal as universal. They also do not establish a differential bias favouring either treatment.³

Concurrent TTM2 participation was addressed through stratification where possible, model adjustment, interaction assessment and an analysis excluding co-enrolled patients. This is more persuasive than ignoring a second intervention. However, absence of a detected interaction in 370 co-enrolled patients does not establish that carbon dioxide and temperature effects are independent. The subgroup analyses similarly provide no validated indication for hypercapnia in patients selected by age, rhythm, shock or resuscitation duration; differences between individual point estimates are not sufficient evidence of treatment-effect heterogeneity.^{1,3}

Finally, safety and applicability are conditional. Patients with intracranial bleeding or oxygen-dependent severe obstructive lung disease were excluded, and profound initial metabolic acidosis delayed treatment. No routine intracranial pressure monitoring was mandated, so the absence of recorded pressure-related treatment cessation does not establish absence of intracranial hypertension. The population was predominantly male, witnessed, initially shockable and treated in high-income-country ICUs. TAME does not settle carbon dioxide management after in-hospital arrest, non-cardiac asphyxial arrest, or when hypercapnia is accepted for another clinical indication.^{1,3}

Body of Evidence

The preceding literature addressed different questions: whether naturally occurring carbon dioxide levels predicted recovery, whether deliberately increasing carbon dioxide changed cerebral physiology, and whether a sustained treatment strategy improved patients' outcomes. These are related steps in developing an intervention, but they do

not provide interchangeable evidence. TAME was designed to test the final clinical proposition rather than to reproduce an association or a physiological response.^{1,2}

The supplied TAME protocol describes a retrospective cohort of 16,542 cardiac arrest admissions to 125 Australian and New Zealand ICUs. Hypocapnia was associated with worse hospital outcomes, whereas hypercapnia was associated with a greater likelihood of discharge home without increased hospital mortality. It also describes a Finnish cohort of 409 patients in which higher early PaCO₂ was associated with favourable neurological outcome at 12 months. These observational findings supported equipoise but could not separate treatment effects from differences in illness, ventilation or recovery that influenced carbon dioxide levels.²

The same protocol describes a physiological crossover study in early cardiac arrest survivors. Deliberately moving from normocapnia to mild hypercapnia increased near-infrared cerebral oxygen saturation, supporting the presence of a modifiable physiological response. Such a study can establish that the measurement responds to the intervention in the studied patients. It cannot determine whether sustained treatment prevents disability or whether greater regional saturation necessarily represents more effective oxygen use by injured brain tissue.²

The CCC phase II trial then randomised 86 patients to 24 hours of PaCO₂ targets of 50–55 or 35–45 mmHg. Biomarkers were assessed in the first 50 patients surviving to day three. Neuron-specific enolase increased less with hypercapnia, with a treatment-by-time interaction P value of 0.04. The corresponding between-group interaction for S100B was not significant. Favourable six-month recovery occurred in 23 patients (59%) versus 18 (46%), and hospital mortality in 11 (26%) versus 15 (37%; P=0.31). The clinical estimates were encouraging but imprecise, while conditioning complete biomarker assessment on survival required caution. The study supported a larger trial rather than clinical adoption.^{2,4}

COMACARE provided a different randomised comparison. This factorial pilot randomised 123 patients, with 120 analysed, to lower-normal or higher-normal PaCO₂ alongside oxygen and blood-pressure allocations for 36 hours. Carbon dioxide targets were 4.5–4.7 versus 5.8–6.0 kPa, using temperature-corrected blood gases. Higher-normal carbon dioxide increased regional cerebral oxygen saturation but did not reduce the primary 48-hour neuron-specific enolase concentration or demonstrate better clinical outcomes. Its small size, narrower carbon dioxide contrast and different temperature correction prevent it from being treated as an earlier definitive test of TAME's intervention. It nevertheless illustrates that a measurable improvement in cerebral oxygenation need not be accompanied by a biomarker or functional benefit.⁵

Taken together, the evidence developed from association through physiological feasibility to a small randomised signal and finally a substantially larger clinical trial. TAME did not reproduce the improvement suggested by the pilot's functional outcomes. That does not make the earlier studies redundant: they justified testing the intervention. It does change their weight in deciding whether to implement it. As of the end of 2023, routine 24-hour targeting of mild hypercapnia lacked demonstrated neurological benefit, while questions about substantially different timing, duration or response-guided selection remained separate research questions rather than supported indications.^{1,2}

Should we target mild hypercapnia after out-of-hospital cardiac arrest?

Not routinely. TAME did not improve six-month neurological outcomes.

References

1. Eastwood G, Nichol AD, Hodgson C, Parke RL, McGuinness S, Nielsen N, et al. Mild Hypercapnia or Normocapnia after Out-of-Hospital Cardiac Arrest. *N Engl J Med* 2023;389(1):45–57.
2. TAME Study Investigators. Protocol, amendments and statistical analysis plans accompanying: Mild Hypercapnia or Normocapnia after Out-of-Hospital Cardiac Arrest. *N Engl J Med* 2023;389(1):45–57.
3. TAME Study Investigators. Supplementary Appendix to: Mild Hypercapnia or Normocapnia after Out-of-Hospital Cardiac Arrest. *N Engl J Med* 2023;389(1):45–57.
4. Eastwood GM, Schneider AG, Suzuki S, Peck L, Young H, Tanaka A, et al. Targeted therapeutic mild hypercapnia after cardiac arrest: A phase II multi-centre randomised controlled trial (the CCC trial). *Resuscitation* 2016;104:83–90.
5. Jakkula P, Reinikainen M, Hästbacka J, Loisa P, Tiainen M, Pettilä V, et al. Targeting two different levels of both arterial carbon dioxide and arterial oxygen after cardiac arrest and resuscitation: a randomised pilot trial. *Intensive Care Med* 2018;44:2112–2121.

STEROHCA

Obling LER, Beske RP, Meyer MAS, Grand J, Wiberg S, Nyholm B, et al. Prehospital high-dose methylprednisolone in resuscitated out-of-hospital cardiac arrest patients (STEROHCA): a randomized clinical trial. Intensive Care Med. 2023;49(12):1467–1478.

Introduction

Restoring spontaneous circulation after cardiac arrest is the beginning of another period of risk. Patients who remain unconscious may develop brain injury, myocardial dysfunction and a systemic inflammatory response despite apparently successful resuscitation. Treatment must therefore address both the cause of the arrest and the consequences of global ischaemia and reperfusion. The difficulty is deciding which elements of this response are amenable to treatment, rather than simply indicators of the damage already sustained. An association between a circulating inflammatory marker and death does not establish that suppressing the marker will improve recovery.¹

Interleukin-6 (IL-6) offers an attractive target within this uncertainty. It participates in the inflammatory response after out-of-hospital cardiac arrest (OHCA), and higher concentrations have been associated with adverse outcomes. Glucocorticoids might interrupt inflammatory injury through several pathways. A large methylprednisolone dose can be administered rapidly, making treatment possible before hospital arrival, when secondary injury may still be evolving. Its effects are not confined to inflammation, however, and the same intervention may alter glucose metabolism and the interpretation of other biochemical measurements.¹

The relevant evidence was incomplete. Earlier trials of steroids combined with vasopressin during in-hospital resuscitation tested a different clinical strategy. Hospital administration of an IL-6 receptor antibody had demonstrated an anti-inflammatory effect without establishing improved neurological recovery. Neither approach answered whether an isolated, early glucocorticoid dose after successful prehospital resuscitation could limit neuronal injury.¹

STEROHCA therefore examined two linked propositions: that prehospital methylprednisolone would reduce systemic inflammation, and that it would reduce biochemical evidence of brain injury. IL-6 and neuron-specific enolase (NSE) were chosen as co-primary outcomes. The trial was an early efficacy study of biological activity, with

longer-term clinical outcomes providing an additional assessment of whether that activity might matter to patients.¹

Synopsis

STEROHCA was an investigator-initiated, randomised, placebo-controlled, phase II superiority trial conducted through the emergency medical services and two cardiac arrest centres in the Capital Region of Denmark. Recruitment ran from 10 October 2020 to 15 July 2022. The Novo Nordisk Foundation and the Research Foundation of Rigshospitalet supported the study; the funders had no role in its design, conduct, analysis, interpretation or writing. An independent physician acting as a trial guardian authorised enrolment, with subsequent confirmation and consent from relatives and patients when possible.¹

Eligible patients were adults with OHCA of suspected cardiac origin who remained unconscious, with a Glasgow Coma Scale score of 8 or less, after at least five minutes of sustained return of spontaneous circulation (ROSC). Important exclusions included an initial rhythm of asystole, more than 30 minutes to ROSC, women of childbearing potential, known treatment limitations, pre-arrest modified Rankin Scale (mRS) scores of 4–5, temperature below 30°C and methylprednisolone allergy. The published protocol additionally required randomisation and treatment initiation within 30 minutes of sustained ROSC. Eligibility could be reassessed during transport or at hospital arrival.^{1,2}

Allocation was 1:1 using a computer-generated sequence in permuted blocks of four. Identical opaque medicine boxes concealed the assignment until opened. The prehospital physician and accompanying assistant then knew the allocation, but did not participate in subsequent hospital care. Patients, hospital personnel and the research team remained blinded. The active treatment was 250 mg intravenous methylprednisolone, prepared from two 125 mg ampoules and administered over five minutes. Placebo was 4 mL isotonic saline given over the same interval. Administration began as soon as practicable after the required five minutes of ROSC and was completed before hospital arrival.¹

Both groups received the centres' post-resuscitation care: temperature management at 36°C in comatose patients, predominantly propofol and fentanyl sedation, and vasopressors or inotropes as required. Comatose patients received prophylactic piperacillin–tazobactam, or cefuroxime under the stated allergy policy, and continuous intravenous insulin for hyperglycaemia. The comparison was a single additional steroid dose against saline within this treatment system.¹

The co-primary outcomes were serial IL-6 and NSE concentrations, sampled at admission and 24, 48 and 72 hours afterwards. Lower concentrations were the anticipated favourable response. The primary analysis included all available measurements; NSE was also examined in patients remaining comatose at hospital arrival. Secondary outcomes included survival and neurological function at discharge and approximately six months, other inflammatory and organ-injury biomarkers, and haemodynamic measurements. Cerebral Performance Category (CPC) scores range from 1 to 5, with higher scores indicating greater impairment and 5 denoting death; mRS scores range from 0 to 6, with 6 denoting death. Discharge assessments used chart review and later assessments used telephone interviews. Registry data established vital status.¹

Sample-size calculations used a single 48-hour measurement because suitable repeated-measurement data were unavailable. Detecting a 20% reduction with 90% power and an alpha of 0.025 required 112 patients for IL-6 and 114 for NSE. The investigators targeted 120 evaluable patients, allowing for early deaths and post-randomisation exclusions through a planned total of 156, or earlier completion if the evaluable target was reached. Recruitment ultimately stopped after 120 patients had survived beyond 72 hours; this was not an efficacy stopping decision. Safety monitoring was overseen by a data and safety monitoring board.¹

Efficacy analyses used a modified intention-to-treat population excluding patients subsequently found ineligible and requiring consent. Log-transformed serial biomarkers were analysed using linear mixed models with unstructured covariance, with estimates back-transformed to geometric means and confidence limits. Maximum-likelihood methods accommodated incomplete measurements. Additional multiple imputation was not performed because the investigators considered the prespecified missingness threshold unmet. The co-primary significance threshold was 0.025; other analyses used 0.05. Survival analyses included an unadjusted Cox model and a model described as predefined, adjusting for sex, age, initial rhythm, time to ROSC and primary percutaneous coronary intervention.¹

The retrospective screening diagram identified 1,976 regional arrests and an estimated 207 potentially eligible patients, of whom 49 were not enrolled for logistical reasons. In total, 158 patients were randomised: 80 to methylprednisolone and 78 to placebo. Twelve and nine, respectively, were excluded afterwards, leaving 68 and 69 in the efficacy analysis. Reasons included non-cardiac aetiology, initial asystole and prolonged resuscitation; single exclusions involved absence of a cardiac arrest diagnosis, childbearing potential and re-arrest before study treatment. All 158 remained in safety follow-up.¹

Median age was approximately 67 years, 81% were men and 94% had an initially shockable rhythm. Median time to ROSC was 18 versus 14 minutes, with more adrenaline and amiodarone administered during resuscitation in the methylprednisolone group. Conversely, cardiogenic shock at admission occurred in 2/68 versus 8/69 patients. Median time from ROSC to randomisation was 20 minutes overall. All patients retained in the efficacy population received the full assigned dose. At least one valid IL-6 measurement was available for 130 patients and an NSE measurement for 129; reported measurement completeness among those surviving beyond 72 hours was 92%.¹

IL-6 trajectories differed between groups, with a treatment-by-time interaction $P < 0.0001$. At admission, estimated geometric mean IL-6 was 15.0 pg/mL in both arms. At 24 hours it was 2.1 pg/mL (95% CI 1.3–3.2) with methylprednisolone versus 29.8 pg/mL (18.9–46.8) with placebo, $P < 0.0001$. Corresponding 48-hour values were 5.7 versus 10.1 pg/mL, $P = 0.04$, and 72-hour values were 4.3 versus 3.4 pg/mL, $P = 0.51$. NSE trajectories did not differ significantly, $P = 0.22$. At 48 hours the geometric means were 18.8 µg/L (95% CI 14.4–24.6) versus 14.8 µg/L (11.2–19.4). The comatose-patient analysis did not change this finding. C-reactive protein trajectories differed, but neurofilament light chain trajectories did not.¹

At 180 days, 51/68 patients (75%) versus 44/69 (64%) were alive. The unadjusted hazard ratio for death was 0.65 (95% CI 0.35–1.20; $P = 0.17$); the adjusted hazard ratio was 0.35 (0.18–0.67; $P = 0.002$). Discharge mortality was 16/68 versus 24/69. Later CPC and mRS distributions did not differ significantly. In the analysis retaining all 158 randomised patients, the unadjusted mortality hazard ratio was 0.80 (95% CI 0.48–1.35).^{1,3}

In the full safety population, 69/80 versus 60/78 patients experienced at least one adverse event and 43/80 versus 44/78 experienced a serious adverse event. Hyperglycaemia was recorded in 30/80 (38%) versus 12/78 (15%), without recorded associated sequelae. These safety denominators differ from the 68 and 69 patients used in the main efficacy tables.^{1,3}

Critique

STEROHCA demonstrates that an early methylprednisolone bolus can substantially alter the systemic inflammatory response after selected OHCA. It does not demonstrate that this treatment protects the brain or improves survival. Those conclusions are compatible: the intervention achieved one biological objective, while the other co-primary outcome and the clinical outcomes did not provide corresponding confirmatory evidence. Describing the trial simply as successful or unsuccessful would obscure what it

contributes. It establishes pharmacological activity and leaves the clinical justification for treatment unresolved.¹

The IL-6 result is persuasive within that limited claim. Random allocation, concealed medicine boxes, blinded hospital treatment and laboratory measurement reduce several opportunities for systematic bias. Delivery of the full dose to retained participants and the marked 24-hour separation make inadequate exposure an unlikely explanation for failure to suppress inflammation. Nevertheless, the strongest separation was transient. The 48-hour P value of 0.04 did not meet the stated co-primary threshold, and the groups converged by 72 hours. A significant overall interaction establishes different trajectories, not sustained superiority at every time point. Repeated treatment might behave differently, but that is another intervention requiring its own evidence.¹

The causal link from peripheral inflammation to meaningful neurological recovery remains unproven. IL-6 was measured in systemic blood, not directly in the brain. Reducing its circulating concentration need not establish that the relevant cerebral inflammatory process has been interrupted. Similarly, NSE is a marker of neuronal injury rather than a complete measure of cognition or functional recovery. Its failure to decrease provides no affirmative support for neuroprotection, but the absence of statistical significance cannot prove that all neurological effects are absent. The concordant lack of a neurofilament light chain signal makes a substantial, readily detectable effect on neuronal injury less convincing.¹

Treatment began early relative to hospital care, but not necessarily early relative to the biological injury. Randomisation occurred a median of 20 minutes after ROSC, following the preceding period of circulatory arrest. Moreover, the first blood sample was obtained after treatment. Similar admission IL-6 concentrations therefore cannot establish equivalence of pre-treatment inflammation or exclude an earlier transient response. The trial cannot determine whether a different timing window, an inflammatory phenotype selected using pretreatment samples, or a longer course would produce a better balance of benefit and harm.¹

The survival findings deserve particular care. The change from an unadjusted hazard ratio of 0.65 to an adjusted value of 0.35 is clinically substantial. Adjustment for strong prognostic variables can improve precision and account for chance imbalance; it is not inherently inappropriate in a randomised trial. Longer resuscitation in the intervention group provides a concrete reason to examine adjustment. However, the model contains several covariates despite only 42 deaths in the efficacy population, calculated from the published 17 and 25 deaths. Primary coronary intervention also occurred after randomisation, so conditioning on it does not automatically retain the interpretation of

the total randomised treatment effect. The adjusted association should not replace the more cautious overall interpretation.¹

Prespecification offers some reassurance, but does not make a small secondary comparison definitive. The published protocol proposed survival adjustment, although its description does not exactly reproduce the final covariate specification. An original-to-final SAP and amendment history were unavailable for establishing the complete chronology. This limits verification of analytic fidelity without establishing that investigators selected their model after viewing treatment effects. Even a fully prespecified adjusted result would remain one secondary finding among many in a study powered for biomarkers, requiring independent clinical confirmation.^{1,2}

Post-randomisation exclusions further weaken a straightforward survival claim. Twenty-one allocated patients were removed from the principal efficacy comparison. Reassessment of an emergency diagnosis is understandable, and most reasons concern characteristics intended to define eligibility. Nevertheless, exclusion after allocation can compromise the protection afforded by randomisation, particularly when treatment has already begun and some staff know the assignment. The full randomised-population survival analysis is therefore valuable. Its hazard ratio of 0.80, with an interval spanning appreciable benefit and harm, supports caution about the magnitude and certainty of the adjusted efficacy-population result.^{1,3}

Missing biomarker measurements present a related but distinct problem. The reported 92% completeness applies to patients surviving beyond 72 hours; it does not describe complete observation of all randomised patients. Death prevents later sampling, while transfer can interrupt biobank collection. A mixed model can use available observations efficiently, but cannot turn the unobserved biomarker course after death into a measured biological outcome. Any assumption that the available data adequately represent the missing trajectories matters when death and injury severity are closely related. The protocol described an additional mortality-related sensitivity analysis whose implementation could not be confirmed in the available results material.^{1,2}

Generalisability is narrow in clinically important ways. Almost all retained patients had a shockable rhythm, resuscitation duration was restricted, and women of childbearing potential and patients with substantial preceding disability were excluded. Some patients regained consciousness before admission. This is not a representative sample of all comatose cardiac arrest survivors, still less all patients undergoing resuscitation. Two specialised centres and physician-led prehospital recruitment provided a workable delivery system; adoption elsewhere would depend on reproducing that early access and

screening. The results cannot establish efficacy in prolonged arrest, asystole, non-cardiac causes or populations excluded for safety reasons.¹

Finally, comparable serious adverse-event totals do not prove safety. Hyperglycaemia occurred more frequently despite protocolised access to intravenous insulin. Universal prophylactic antibiotics in comatose patients also define the setting in which infections were observed. A single modest-sized trial cannot exclude uncommon harms or establish the safety of larger or repeated doses. STEROHCA provides a sound reason to continue investigating an early anti-inflammatory strategy, but insufficient evidence to make a 250 mg methylprednisolone bolus part of routine post-arrest care. A larger trial should determine whether any survival or functional gain outweighs these treatment burdens.¹

Body of Evidence

The preceding literature addresses several different questions under the broad heading of steroids in cardiac arrest. Giving a drug during cardiopulmonary resuscitation aims partly to restore circulation; treating established post-arrest shock aims to reduce circulatory failure; suppressing inflammation soon after ROSC aims to prevent subsequent injury. These strategies differ in their populations, timing and accompanying treatments. Their results can inform the rationale for STEROHCA without functioning as direct replications of its intervention.

The 2013 Greek trial by Mentzelopoulos and colleagues randomised 268 patients with in-hospital cardiac arrest requiring adrenaline. The experimental regimen combined vasopressin during resuscitation with 40 mg methylprednisolone, followed by hydrocortisone for post-resuscitation shock; controls received adrenaline and matched placebo. Sustained ROSC and survival to discharge with CPC 1–2 were co-primary outcomes. Favourable neurological survival occurred in 18/130 versus 7/138 patients, with an odds ratio of 3.28 (95% CI 1.17–9.20). This supported further investigation of the combined regimen. It could not identify which component was responsible, or establish the efficacy of a larger, isolated steroid dose after an out-of-hospital arrest.⁴

Donnino and colleagues subsequently studied a more direct post-arrest steroid strategy in 50 patients requiring vasopressors for at least one hour; 76% had OHCA. Patients received hydrocortisone 100 mg every eight hours or placebo, for up to seven days with discontinuation after shock resolution. The primary endpoint was time to shock reversal, analysed with death as a competing event. The hazard ratio was 0.83 (95% CI 0.40–1.75; $P=0.63$), and survival to discharge was 7/25 versus 9/25. The imprecision limits exclusion of benefit. Nevertheless, this study offered no established clinical advantage from

treating post-arrest shock with that regimen, and its indication and exposure differ substantially from STEROHCA's prehospital bolus.⁵

IMICA examined inflammation more directly. In this blinded Danish trial, 80 comatose OHCA patients in the analysed population received the IL-6 receptor antibody tocilizumab or placebo in hospital. The primary endpoint was the C-reactive protein response over 72 hours. At 24 hours, tocilizumab reduced C-reactive protein by 84% relative to placebo (95% CI 76–90% reduction); myocardial-injury markers also decreased, without a demonstrated survival or neurological benefit. This is useful mechanistic context for STEROHCA: inflammatory suppression was achievable through a different pathway, but its translation into better recovery remained unestablished. Tocilizumab and methylprednisolone should not be assumed to have interchangeable biological effects.⁶

The 2021 Danish VAM-IHCA trial provided a larger test of treatment during resuscitation. Its analysis included 501 patients allocated to vasopressin and 40 mg methylprednisolone or placebo after adrenaline administration. ROSC, the primary outcome, occurred in 100/237 versus 86/264 patients: risk ratio 1.30 (95% CI 1.03–1.63). Thirty-day survival was 23/237 versus 31/264, with a risk ratio of 0.83 (0.50–1.37), and favourable neurological outcome was 7.6% in both groups. Thus improved initial resuscitation did not establish improved longer-term recovery. As in the earlier combination trial, the steroid's independent contribution could not be isolated; the intervention also began before ROSC, unlike STEROHCA.⁷

Across these trials, the endpoints closest to the intervention may respond while the outcomes most important to patients remain uncertain. That pattern is not proof that the underlying mechanisms are irrelevant. It means that neither a favourable laboratory response nor more frequent ROSC can substitute for a sufficiently precise assessment of survival and neurological function. Differences in baseline injury, timing, co-treatment and dose also prevent attributing disagreements simply to the presence or absence of a steroid.

By the end of 2023, STEROHCA added evidence that a brief prehospital glucocorticoid intervention could suppress systemic inflammation after successful resuscitation. It did not supply the missing confirmation of clinical benefit. Its appropriate place in this sequence is as a mechanistic trial that justifies a more definitive patient-centred test, while discouraging routine adoption based solely on the IL-6 response or the favourable adjusted survival estimate.¹

Should we routinely give 250 mg methylprednisolone after prehospital resuscitation from cardiac arrest?

Not routinely. STEROHCA demonstrates transient inflammatory suppression, but does not establish neurological or survival benefit.

References

1. Obling LER, Beske RP, Meyer MAS, Grand J, Wiberg S, Nyholm B, et al. Prehospital high-dose methylprednisolone in resuscitated out-of-hospital cardiac arrest patients (STEROHCA): a randomized clinical trial. *Intensive Care Med.* 2023;49(12):1467–1478.
2. Obling LER, Beske RP, Wiberg S, Folke F, Moeller JE, Kjaergaard J, et al. Steroid treatment as anti-inflammatory and neuroprotective agent following out-of-hospital cardiac arrest: a randomized clinical trial. *Trials.* 2022;23:952.
3. Obling LER, Beske RP, Meyer MAS, Grand J, Wiberg S, Nyholm B, et al. Electronic supplementary material to: Prehospital high-dose methylprednisolone in resuscitated out-of-hospital cardiac arrest patients (STEROHCA): a randomized clinical trial. *Intensive Care Med.* 2023;49:1467–1478.
4. Mentzelopoulos SD, Malachias S, Chamos C, Konstantopoulos D, Ntaidou T, Papastilianou A, et al. Vasopressin, steroids, and epinephrine and neurologically favorable survival after in-hospital cardiac arrest: a randomized clinical trial. *JAMA.* 2013;310(3):270–279.
5. Donnino MW, Andersen LW, Berg KM, Chase M, Sherwin R, Smithline H, et al. Corticosteroid therapy in refractory shock following cardiac arrest: a randomized, double-blind, placebo-controlled, trial. *Crit Care.* 2016;20:82.
6. Meyer MAS, Wiberg S, Grand J, Meyer ASP, Obling LER, Frydland M, et al. Treatment effects of interleukin-6 receptor antibodies for modulating the systemic inflammatory response after out-of-hospital cardiac arrest (the IMICA trial): a double-blinded, placebo-controlled, single-center, randomized, clinical trial. *Circulation.* 2021;143(19):1841–1851.
7. Andersen LW, Isbye D, Kjaergaard J, Kristensen CM, Darling S, Zwisler ST, et al. Effect of vasopressin and methylprednisolone vs placebo on return of spontaneous circulation in patients with in-hospital cardiac arrest: a randomized clinical trial. *JAMA.* 2021;326(16):1586–1594.

ARREST

Patterson T, Perkins GD, Perkins A, Clayton T, Evans R, Dodd M, et al. Expedited transfer to a cardiac arrest centre for non-ST-elevation out-of-hospital cardiac arrest (ARREST): a UK prospective, multicentre, parallel, randomised clinical trial. Lancet. 2023;402(10410):1329–1337.

Introduction

After return of spontaneous circulation following out-of-hospital cardiac arrest (OHCA), the ambulance team must decide where to take the patient. A specialist cardiac arrest centre offers immediate cardiology assessment, coronary intervention, advanced circulatory support and experienced critical care. The nearest emergency department offers earlier arrival, broad diagnostic assessment and access to local intensive care. The decision concerns the whole pathway: additional travel, the point of hospital entry and the treatments that subsequently become available. A hospital's designation alone cannot establish which pathway improves recovery.^{1,2}

The question is particularly difficult when the post-resuscitation electrocardiogram does not show ST-segment elevation. A coronary cause remains possible, but so do primary arrhythmia, cardiomyopathy and non-cardiac disease. A prehospital impression of cardiac aetiology is necessarily provisional. Direct delivery to a catheter laboratory could shorten access to reperfusion for some patients while offering little immediate advantage to others. Whether this selection can be made reliably at the scene is central to the value of a universal specialist-destination policy.¹

Observational comparisons had encouraged regionalisation. In London's pre-trial audit, survival appeared substantially higher at cardiac arrest centres. Yet those centres also received a selected population, and patients' prognosis could influence their destination. Differences in patient selection, available treatment and the likelihood of transfer make those associations difficult to interpret causally. A randomised comparison was needed before committing all patients without ST elevation to the more specialised pathway.²

ARREST tested that policy directly within the London ambulance and hospital network. Its intervention was expedited delivery to a specialist centre's catheter laboratory, compared with delivery to the geographically closest emergency department. By allocating the destination before hospital arrival, the trial addressed a decision that ambulance services actually face, while leaving subsequent treatment responsive to the diagnosis and the patient's condition.¹

Synopsis

ARREST was an investigator-led, prospective, open-label, parallel-group superiority trial conducted by the London Ambulance Service and 35 receiving hospitals. Seven were cardiac arrest centres with emergency interventional cardiology, cardiac surgery and specialist intensive care facilities. Four of these also had emergency departments; the usual-care pathway could therefore lead to one of 32 emergency departments across the network. The British Heart Foundation funded the trial without involvement in design, data collection, analysis, interpretation or writing. Enrolment ran from 15 January 2018 to 1 December 2022, with two prolonged pauses during the COVID-19 pandemic.^{1,2}

Adults with OHCA and return of spontaneous circulation were eligible if paramedics considered the cause cardiac and the post-resuscitation ECG did not meet ST-elevation myocardial infarction criteria. Exclusions included a do-not-attempt-resuscitation order, suspected pregnancy, presumed non-cardiac cause, significant trauma and arrest occurring after an existing care pathway had been set during transport. Neither a shockable initial rhythm nor a specified coma score was required. Eligibility was assessed at the scene without investigator oversight. Prior consent was waived for the emergency intervention; subsequent patient consent or a personal or professional consultee process governed continued participation and data use.^{1,2}

Attending paramedics contacted an advanced paramedic practitioner dispatch desk, where designated staff used a secure web system to allocate patients 1:1. Randomisation used permuted blocks of four and six without stratification. The intervention could not be masked to ambulance staff, patients or hospital clinicians, and follow-up staff were also unblinded. The SAP specifies that outcomes were not independently adjudicated.^{1,3}

For intervention patients, the receiving cardiac arrest centre was pre-alerted and the patient was taken directly to its catheter laboratory, with cardiology and intensive care assessment available around the clock. The centres could provide ventilation, haemodynamic support, temperature control, diagnostic imaging, immediate reperfusion and mechanical circulatory support. Angiography or any particular organ-support treatment was not mandated. Standard-care patients went to the geographically closest emergency department and then received local hospital care, including subsequent specialist referral when indicated. Arrival at the assigned destination completed the trial intervention. Both groups were required to defer neuroprognostication until at least 72 hours after arrest before withdrawal of life-sustaining treatment.^{1,2}

The primary endpoint was all-cause mortality 30 days after randomisation. Secondary outcomes included mortality at three months, neurological function at discharge and three months, and EQ-5D-5L health status at discharge. The modified Rankin Scale (mRS), from 0 for no symptoms to 6 for death, was the primary neurological measure; Cerebral Performance Category (CPC), from 1 to 5 with death at 5, was also assessed. Favourable status meant mRS 0–3 or CPC 1–2 for the respective scale. Six- and twelve-month mortality and economic analyses were planned but were not complete in this report.^{1,3}

The sample-size calculation assumed 60% control mortality and sought a reduction to 50%. A total of 860 participants provided 80% power at a 5% significance level, allowing up to 10% loss to follow-up. The anticipated treatment effect reflected the combined potential of specialist interventions and earlier observational estimates. An independent data and safety monitoring committee reviewed accumulating data; recruitment reached the planned sample rather than ending for a demonstrated efficacy or futility boundary.^{1,3}

The primary comparison used participants in their allocated groups after removal of withdrawals, with analysis restricted to known 30-day mortality. Generalised linear models with log and identity links estimated risk ratios and risk differences. A supplementary adjusted model included age, sex, initial rhythm, witnessed status, bystander CPR, arrest-to-ROSC time and arrest location. Missing baseline covariates were multiply imputed; logistic regression produced an adjusted odds ratio when the risk-ratio model failed to converge. Ordered logistic regression assessed neurological distributions. Secondary analyses used available outcomes, and the SAP described additional missing-score and survivor-only analyses. Secondary and subgroup findings were exploratory.^{1,3}

Of 4,019 patients assessed, 862 were randomised, 431 to each group. Seventeen intervention and eighteen control patients were withdrawn, leaving the report's intention-to-treat populations of 414 and 413. A further three and one had unknown 30-day mortality, leaving 411 and 412 for the primary analysis. Reasons for withdrawal included declined consent, overseas status or discharge abroad, and two intervention allocations made in error. Nineteen protocol deviations were recorded, including three pathway crossovers. Later inpatient transfer to a cardiac arrest centre occurred in 70/406 standard-care patients and was permitted by the protocol.^{1,2}

Mean age was approximately 63–64 years, 68% of participants with recorded sex were male, and 55% had a shockable initial rhythm. Median arrest-to-ROSC time was 24 versus 25 minutes among those with recorded times. Arrest-to-hospital arrival was 84 minutes (IQR 68–104; n=332) versus 77 minutes (63–96; n=328), with a reported difference of

seven minutes (95% CI 2–12). A cardiac cause was ultimately identified in 260/414 versus 245/413 patients; the cause remained unknown in 68 and 89, respectively.¹

Treatment differed after arrival. Coronary angiography was performed in 231/412 patients (56%) versus 153/410 (37%); among those with timing data, median arrest-to-angiography time was 2.3 versus 5.7 hours. Intensive care admission occurred in 330/414 (80%) versus 286/413 (69%). Ventilatory support was used in 353/412 versus 312/410, haemodynamic support in 297/412 versus 252/406, and renal support in 46/411 versus 34/403. Cardiogenic shock was identified in 112/408 versus 93/406. Mechanical circulatory support was uncommon in both groups.^{1,2}

At 30 days, 258/411 intervention patients and 258/412 controls had died, approximately 63% in each group. The mortality table gives an unadjusted risk ratio of 1.00 (95% CI 0.90–1.11; $P=0.96$) and an absolute risk difference of 0.2 percentage points (95% CI –6.5 to 6.8). The adjusted odds ratio was 1.09 (0.73–1.63). Three-month mortality was 267/411 versus 263/411, risk ratio 1.02 (0.92–1.12). Neither the discharge mRS distribution, proportional odds ratio 1.00 (0.76–1.32), nor the three-month distribution, 0.98 (0.73–1.31), showed a clear treatment effect.¹

Favourable mRS status at discharge occurred in 130/413 versus 130/402 patients; at three months it occurred in 119/399 versus 119/390. EQ-5D-5L index values were available for only 97 and 92 patients, with a mean difference of –0.04 (95% CI –0.12 to 0.05). The age subgroup interaction was significant, with opposite treatment directions in patients below 57 years and those aged 57–71. Eight of 414 versus three of 413 patients experienced separately reported serious adverse events, none judged intervention-related. Expected complications of cardiac arrest were captured separately, rather than counted uniformly within this small serious-event total.^{1,2}

Critique

ARREST provides a credible answer to a specific systems question. Within this London network, routinely bypassing the nearest emergency department for direct catheter-laboratory delivery did not produce the large survival gain anticipated. Concealed central allocation before hospital arrival directly addresses the selection bias that limits observational comparisons of hospital destinations. The objective mortality endpoint, recruitment across an entire urban network and very few initial pathway crossovers strengthen that conclusion. The trial's value lies in testing an apparently attractive service policy under conditions in which the policy could realistically operate.^{1,2}

The result should not be expanded into equivalence between all hospitals or all models of specialist care. The 95% confidence interval for the absolute mortality difference extends from a 6.5-percentage-point reduction to a 6.8-point increase. It excludes the targeted ten-point benefit in the analysed population but still permits differences that patients and services could consider important. A superiority trial with this precision cannot establish that the two pathways have identical effects. The appropriate conclusion is that a broad mandatory diversion policy lacks demonstrated survival benefit in this setting, rather than that specialist expertise has no value.¹

The comparator is essential to that interpretation. Standard care meant entry through the nearest emergency department, not exclusion from specialist treatment. Four emergency departments were themselves within cardiac arrest centres, and subsequent inpatient transfer was permitted and used. The 70 later transfers should therefore not be described as 70 failures of randomisation. They form part of a strategy in which initial stabilisation and diagnostic assessment guide referral. The trial estimates the effect of immediate universal specialist routing compared with this existing pathway; it does not compare lifelong separation from specialist care with guaranteed access to it.^{1,2}

There was meaningful treatment separation despite that overlap. More intervention patients underwent angiography, and angiography occurred earlier. They were also more often admitted to intensive care and received more organ support. A completely ineffective implementation of the assigned pathway is therefore an inadequate explanation for the neutral mortality result. However, assignment changed several things simultaneously: travel, receiving team, hospital entrance, diagnostic emphasis and subsequent treatment. ARREST cannot identify the independent effect of bypassing the emergency department, of angiography, or of the additional support delivered.¹

The broad prehospital eligibility criteria improve relevance to ambulance decision-making while reducing biological specificity. A cardiac cause was confirmed in only about three-fifths of patients, despite the initial cardiac impression. This is not simply poor recruitment: uncertain diagnosis is intrinsic to the decision being tested. A policy requiring every presumed cardiac arrest survivor without ST elevation to enter a catheter laboratory must work despite that uncertainty. Conversely, failure of the broad policy does not rule out benefit for a more accurately identifiable subgroup requiring an urgent specialist procedure. Such a subgroup would need a usable prospective selection rule and independent evidence.^{1,2}

The observed seven-minute difference was measured from arrest to hospital arrival. It should not be mistaken for a fully observed, isolated ambulance travel interval, nor extrapolated directly to rural transport. Timing data were incomplete, and the contrast

incorporates the entire prehospital course. The longer interval is a plausible disadvantage of diversion, but the trial did not separately randomise travel time. Similarly, greater cardiogenic shock recognition and organ support after specialist arrival could reflect illness, clinical thresholds, access or treatment consequences. These post-allocation observations cannot establish that diversion caused circulatory deterioration.¹

Open treatment assignment was unavoidable, but it remains relevant to care decisions and outcome assessment. The common requirement for delayed neuroprognostication was a useful safeguard; it does not demonstrate identical withdrawal practices across all hospitals. Neurological scores were assessed without masking or independent adjudication, and their completeness was less secure than vital status. The SAP acknowledges that mRS collection began after the earliest recruits and that obtaining complete functional scores was difficult. These features limit confidence in small functional differences and in the apparent absence of differences among patients with missing assessments.^{1,3}

The reported intention-to-treat population also falls short of retaining every randomised patient. Thirty-five withdrawals preceded that population's definition, and four additional participants lacked the primary endpoint. Losses were modest and nearly balanced, which is reassuring, but balance of numbers does not prove balance of prognosis. Missing baseline covariates were imputed for adjustment; missing primary outcomes were not made known by that procedure. The adjusted odds ratio supports the primary interpretation but is a different effect measure from the unadjusted risk ratio and should not be compared numerically as though the two were interchangeable.^{1,3}

The age finding is insufficient to set a destination threshold. The SAP planned age tertiles, and the interaction P value of 0.0029 warrants attention. However, the apparent advantage below 57 years reversed in the middle tertile and did not follow a consistent progression across age. Multiple subgroup questions were examined, and the particular age boundaries arose from the sample's distribution. A selective policy based on those cutpoints would give an exploratory result more authority than the overall evidence supports. Survivor-only neurological comparisons likewise condition on an event occurring after allocation and cannot establish a randomised neurological harm effect.^{1,2,3}

The SAP is dated 28 April 2023, after recruitment and the principal three-month follow-up had ended. It states that approval must precede database lock and unblinded final analysis, and that the primary objective had not changed. The available documents

do not independently establish that chronology or provide the full earlier amendment history. Several planned supplementary analyses are not displayed in the supplied report. These are limits to verification, rather than evidence of selective analysis. They reinforce the priority of the clearly defined primary endpoint over isolated secondary findings.³

Finally, the serious-event tally is not a complete measure of treatment burden: expected arrest complications followed separate reporting rules. Greater resource use without a demonstrated survival gain raises a service-planning question, but the unfinished economic analyses cannot establish cost-effectiveness. Generalisability also depends on capable local hospitals, short urban distances and an established referral network. ARREST supports retaining local emergency assessment as a reasonable default in comparable systems, with specialist transfer directed by clinical need. It cannot determine the best pathway where the nearest hospital lacks effective post-arrest care.^{1,2}

Body of Evidence

The evidence leading to ARREST combines feasibility research, observational hospital comparisons and trials of specific treatments available at specialist centres. These strands answer different questions. A centre can be associated with better outcomes because it receives patients more likely to recover; an intervention can be available there without improving outcomes when applied routinely. Neither observation establishes the effect of changing every eligible patient's prehospital destination. That distinction explains why a trial of the whole pathway remained necessary.¹

The 2017 ARREST pilot directly addressed feasibility. Adults with witnessed ventricular-fibrillation OHCA of presumed cardiac origin, without ST elevation, were allocated to expedited specialist transfer or the closest hospital. Forty of 63 eligible patients were randomised from 118 screened cases, and primary and secondary outcome collection was achieved in 83%. Thirty-day mortality among those with data was 9/18 versus 6/15. These small numbers did not establish efficacy. The pilot showed that emergency randomisation and pathway delivery were practicable, while identifying follow-up as an operational concern. Its restricted witnessed, shockable-rhythm population also differed from the broader definitive trial.⁴

London's 2017–18 audit provided a striking observational rationale. As reproduced in the ARREST supplement, the mean centre-level survival figures were 45.7% at cardiac arrest centres and 14.4% at other hospitals. Specialist centres received approximately twice the median annual arrest volume. However, this audit included both ST-elevation and

non-ST-elevation arrests and all presumed causes, with survival calculated among patients whose outcomes were known. Destination was selected clinically. The large difference therefore combined patient selection, hospital characteristics and possibly treatment effects; it could not predict the effect of randomising the narrower ARREST population. Its attenuation under random allocation is a substantive finding, not merely a disappointing failure to reproduce an expected benefit.²

COACT, published in 2019, tested one proposed advantage of specialist arrival: immediate coronary angiography. It randomised 552 unconscious OHCA survivors with a shockable rhythm, without ST elevation, shock or an obvious non-coronary cause, to immediate angiography or angiography after neurological recovery. Urgent angiography remained available for deterioration. Among the 538 patients analysed, 90-day survival was 176/273 versus 178/265, odds ratio 0.89 (95% CI 0.62–1.27). This did not demonstrate a survival advantage from routine immediate angiography in that selected population. It informed ARREST's rationale but could not replace a destination trial, because both COACT groups were already in hospital and only the timing strategy was allocated.⁵

The 2021 TOMAHAWK trial extended that question. It randomised 554 resuscitated patients without ST elevation to immediate angiography or initial intensive care assessment with delayed or selective angiography. The primary analysis included 530 patients. Thirty-day mortality was 143/265 versus 122/265, hazard ratio 1.28 (95% CI 1.00–1.63; $P=0.06$). The result provided no support for routine immediate angiography and did not establish a precise harm effect. This distinction matters when interpreting ARREST: earlier access to a catheter laboratory is an intermediate change in care, and its clinical value depends on who needs the procedure, what other care is delayed and the eventual outcome.⁶

The accompanying 2023 commentary by Hansen, Folke and Granger placed the destination question alongside the importance of established resuscitation care and allowing adequate time for neurological recovery. That is a useful clinical perspective, but commentary does not add randomised patient data. The service conclusion should remain anchored to ARREST's own comparison and its local-hospital capabilities, rather than a general preference either for more technology or for less specialised care.⁷

By the end of 2023, ARREST supplied the missing direct randomised test of routine specialist routing for this non-ST-elevation population. Together with COACT and TOMAHAWK, it challenged the assumption that more immediate invasive access necessarily improves recovery. It supported a pathway combining prompt emergency assessment, competent post-arrest care and selective specialist referral in a network

resembling London's. It left open the value of different receiving arrangements, better selection for urgent specialist intervention and systems with markedly different transport distances or local resources.¹

Should every resuscitated OHCA patient without ST elevation bypass the nearest emergency department for a cardiac arrest centre?

No. ARREST does not support routine diversion to a cardiac arrest centre's catheter laboratory in a system with capable local hospitals and specialist referral pathways.

References

1. Patterson T, Perkins GD, Perkins A, Clayton T, Evans R, Dodd M, et al. Expedited transfer to a cardiac arrest centre for non-ST-elevation out-of-hospital cardiac arrest (ARREST): a UK prospective, multicentre, parallel, randomised clinical trial. *Lancet*. 2023;402(10410):1329–1337.
2. Patterson T, Perkins GD, Perkins A, Clayton T, Evans R, Dodd M, et al. Supplementary appendix to: Expedited transfer to a cardiac arrest centre for non-ST-elevation out-of-hospital cardiac arrest (ARREST). *Lancet*. 2023;402:1329–1337.
3. ARREST Trial Investigators. Statistical analysis plan, version 1.0, 28 April 2023. In: Supplementary appendix to the ARREST trial. *Lancet*. 2023;402:1329–1337.
4. Patterson T, Perkins GD, Joseph J, Wilson K, Van Dyck L, Robertson S, et al. A Randomised trial of Expedited transfer to a cardiac arrest centre for non-ST elevation ventricular fibrillation out-of-hospital cardiac arrest: the ARREST pilot randomised trial. *Resuscitation*. 2017;115:185–191.
5. Lemkes JS, Janssens GN, van der Hoeven NW, Jewbali LSD, Dubois EA, Meuwissen M, et al. Coronary angiography after cardiac arrest without ST-segment elevation. *N Engl J Med*. 2019;380(15):1397–1407.
6. Desch S, Freund A, Akin I, Behnes M, Preusch MR, Zelniker TA, et al. Angiography after out-of-hospital cardiac arrest without ST-segment elevation. *N Engl J Med*. 2021;385(27):2544–2553.
7. Hansen CM, Folke F, Granger CB. Back to basics for out-of-hospital cardiac arrest. *Lancet*. 2023;402(10410):1300–1301.

Possible Reporting Inconsistencies. The treatment columns for three-month modified Rankin scores among survivors appear reversed in supplementary Table 8: the cardiac-arrest-centre counts reproduce the standard-care counts in main Table 3, and vice versa. The supplementary ordinal-effect column is also labelled "risk ratio", although the SAP specifies an ordered logistic model, which estimates an odds ratio. These issues concern neurological-outcome reporting; the primary mortality counts agree.^{1,2,3}

INCEPTION

Suverein MM, Delnoij TSR, Lorusso R, Brandon Bravo Bruinsma GJ, Otterspoor L, Elzo Kraemer CV, et al. Early Extracorporeal CPR for Refractory Out-of-Hospital Cardiac Arrest. N Engl J Med. 2023;388(4):299–309.

Introduction

Extracorporeal cardiopulmonary resuscitation (ECPR) uses venoarterial extracorporeal membrane oxygenation (ECMO) to restore perfusion while clinicians identify and treat the cause of an otherwise refractory cardiac arrest. Venous blood is drained through an extracorporeal circuit, oxygenated and returned to the arterial circulation. For a patient with persistent ventricular fibrillation and a potentially treatable coronary lesion, this can create an opportunity for definitive treatment. The clinical objective, however, is survival with acceptable neurological function. Establishing extracorporeal blood flow cannot by itself demonstrate that the brain has survived the preceding arrest.¹

ECPR also requires several time-sensitive decisions to work together. Ambulance clinicians must recognise a suitable patient, continue effective resuscitation and arrange transport. A receiving team must prepare, reassess eligibility and obtain vascular access while conventional cardiopulmonary resuscitation (CPR) continues. Subsequent critical care must manage the underlying disease, organ injury and complications of mechanical support. Failure at any of these stages can reduce the benefit of the overall strategy, even when the physiological rationale for restoring circulation remains persuasive. Conversely, patients who regain spontaneous circulation while the team prepares may survive without ever needing the circuit.²

INCEPTION tested whether offering ECPR within a multicentre Dutch system improved neurological survival over continued conventional resuscitation. Its importance lies in the transition from a promising rescue procedure to a service that several hospitals and ambulance systems could deliver. The study also confronts a practical problem in emergency research: randomisation may be needed before complete information about eligibility is available. The resulting trial must be interpreted through its actual recruitment, treatment delivery and follow-up, as well as its primary effect estimate. For clinicians considering an ECPR programme, these details determine what the trial can reasonably say about benefit in their own setting.^{1,3}

Synopsis

INCEPTION was an investigator-initiated, pragmatic, open-label, randomised trial involving ten cardiosurgical centres and twelve emergency medical services in the Netherlands. Recruitment ran from May 2017 to February 2021. The participating regions included urban and rural areas and covered approximately 8.5 million people. Recruitment expanded beyond the original centres and was interrupted during the COVID-19 pandemic. Funding came from the Netherlands Organisation for Health Research and Development and Maquet Cardiopulmonary, part of Getinge; the funders had no role in trial design, conduct, analysis or reporting.^{1,3}

Eligible patients were aged 18–70 years, had a witnessed out-of-hospital cardiac arrest with an initial shockable ventricular rhythm, and had not regained spontaneous circulation after 15 minutes of advanced life support. Bystander basic life support was required, with an exception for arrests witnessed by emergency medical services when advanced life support began immediately. Exclusions included sustained haemodynamic recovery within 15 minutes, advanced heart or pulmonary disease, disseminated malignancy, pregnancy, severe pre-existing neurological disability, major trauma, relevant treatment limitations, bilateral femoral bypass surgery and other contraindications to ECPR. An expected interval exceeding 60 minutes from arrest to the start of cannulation also excluded participation.^{1,3}

Patients were allocated equally using a web-based randomisation system, with variable blocks of two, four and six and stratification by centre. Hospital teams could randomise during transport to allow preparation before arrival; ambulance personnel remained unaware of allocation. Eligibility was checked again at hospital admission. Patients subsequently found ineligible could be excluded, but an actual cannulation start beyond 60 minutes did not itself remove a patient from the principal analysis. Stable spontaneous circulation before ECPR initiation meant that the procedure was unnecessary, while the patient remained in the assigned group. Deferred consent procedures included patients, representatives and a waiver for patients who died before consent could be obtained.^{1,2,4}

The intervention added extracorporeal support to conventional advanced life support. Teams used a Cardiohelp system with HLS advanced circuitry and femoral arterial and venous cannulation; the cannulation approach and subsequent treatment followed local practice. Conventional resuscitation continued in the control group, with hospital transport during ongoing CPR available in both groups. Coronary intervention and post-resuscitation critical care were guided by clinical need and local protocols. The original and final protocols described temperature control and multimodal neurological

assessment, but the trial did not establish that a single post-resuscitation regimen was delivered uniformly across centres.^{1,2}

The primary endpoint was survival at 30 days with a Cerebral Performance Category (CPC) score of 1 or 2. On this five-category scale, these categories indicate good cerebral performance or moderate disability compatible with independent daily activities; higher categories represent greater impairment or death. An independent neurologist unaware of treatment assignment assessed the endpoint by telephone or in person. Secondary outcomes included neurological survival at later follow-up, resuscitation duration, hospital and intensive care unit (ICU) survival and stay, organ support, reasons for treatment discontinuation and serious adverse events.^{1,4}

The initial calculation assumed favourable neurological survival would rise from 8% to 30%. With 80% power and two-sided alpha of 0.05, 49 patients per group were required, increased to 55 per group for anticipated discontinuation. An interim analysis after 40 randomisations did not lead to an effect-based increase. Subsequent recognition that some intervention patients regained spontaneous circulation before ECPR led to a revised target of 134, intended to retain the anticipated number undergoing the procedure. The final primary analysis used a centre-adjusted logistic mixed model, with small centres grouped; the final significance threshold was 0.048 after allowance for interim testing. Secondary confidence intervals were not adjusted for multiplicity.^{1,4}

Of 160 randomised patients, 26 were excluded after reassessment, leaving 70 allocated to ECPR and 64 to conventional CPR. Reasons included an unwitnessed arrest or unknown delay in nine, age above 70 in five, a non-shockable rhythm in five, early return of spontaneous circulation in three and other reasons in four. Mean ages were 54 and 57 years, and approximately nine in ten participants were men. CPR had begun within five minutes in 69 of 70 intervention patients and 61 of 64 controls. Acute myocardial infarction was the attributed cause in 51 and 52 patients, respectively.¹

ECPR was started in 52 of 70 intervention patients. Of the 18 who did not undergo the procedure, 13 regained stable spontaneous circulation, three encountered logistical failure and two had further treatment judged futile. Extracorporeal circulation was successfully established in 46 of the 52 attempts; five failures were procedural and one involved inadequate flow. Three control patients crossed over to ECPR, and all subsequently died. In intervention patients with available measurements, cannulation began a mean of 58 minutes after arrest. The median interval from hospital arrival to cannulation start was 16 minutes, cannulation duration was 20 minutes, and arrest-to-extracorporeal-flow time was 74 minutes, with an interquartile range of 63–87 minutes.^{1,3}

At 30 days, favourable neurological survival occurred in 14 of 70 patients assigned to ECPR (20%) and 10 of 62 evaluable controls (16%): adjusted odds ratio 1.4, 95% confidence interval 0.5–3.5; $P=0.52$. One control patient lacked the independent neurological assessment and another declined further participation. At six months, the corresponding counts were 14 of 70 and 10 of 63, with an odds ratio of 1.3 (95% confidence interval 0.5–3.3). Survival to hospital discharge was 14 of 70 versus 13 of 64. ICU admission occurred in 57 of 70 versus 23 of 64, and percutaneous coronary intervention in 34 versus 14 patients.¹

The mean number of serious adverse events per patient was 1.4 in the ECPR group and 1.0 in controls. More than one event occurred in 33 of 70 versus 14 of 64 patients. Major bleeding affected 11 versus two patients, limb ischaemia four versus none, and failure to maintain ECMO flow nine versus none. These outcomes arose against very different early mortality: 13 intervention patients and 41 controls died before ICU admission. The primary comparison did not demonstrate superiority, and its confidence interval left considerable uncertainty about benefit or harm.^{1,3}

Critique

INCEPTION addressed a clinically relevant strategy and used a meaningful primary endpoint. Random allocation, centre stratification, prospective specification of neurological survival and an independent masked assessor are substantial strengths. The intervention was tested across several receiving hospitals within an existing ambulance system, making the study informative about implementation beyond one exceptionally experienced service. Hospital teams could not be blinded, but keeping ambulance personnel unaware of assignment reduced the opportunity for allocation to alter prehospital treatment. Almost complete primary follow-up among the retained trial population also limits concern about attrition after enrolment.^{1,2,4}

The main estimate is nevertheless too imprecise to establish comparable efficacy. An odds ratio of 1.4 with limits of 0.5 and 3.5 includes both appreciable harm and substantial benefit. The trial was designed around a large absolute improvement of 22 percentage points, while favourable survival in controls was twice the assumed 8%. Increasing recruitment to compensate for patients not receiving ECPR did not guarantee adequate precision for a smaller, clinically worthwhile strategy effect. The absence of statistical significance therefore answers whether superiority was demonstrated, without settling whether the procedure has useful efficacy in an optimally selected patient.¹

The description of the primary analysis as intention to treat requires qualification. Twenty-six patients were removed after randomisation, so the published comparison

concerns a modified population. Reassessment during an emergency is understandable: age, the initial rhythm and whether the arrest was witnessed may be uncertain during transport. However, exclusions determined after treatment assignment can weaken the balance created by randomisation. The analysis plan anticipated exclusions and required review by steering-committee members, but the principal flow diagram does not provide the original allocation of the 26 excluded patients. Their absence cannot be assumed harmless simply because the remaining baseline characteristics appear similar.^{1,4}

Treatment receipt further defines the question being answered. Only 46 of 70 intervention patients had extracorporeal circulation successfully established. Some of this shortfall reflected logistical or technical failure, but 13 patients recovered spontaneous circulation without requiring the planned procedure. These are clinically different events. Regaining circulation is a desirable pathway within the assigned strategy, whereas inability to provide necessary support is a delivery failure. Removing all patients who never received ECMO would mix these pathways and select groups using events that occurred after randomisation. The trial estimates the effect of offering ECPR under its delivery conditions, rather than the isolated effect of successful cannulation.^{1,3}

Timing is central to that interpretation. Eligibility used an expected cannulation start within 60 minutes; this was not a requirement to establish extracorporeal flow within 60 minutes. The observed median arrest-to-flow interval was 74 minutes, and flow-time data were available for 44 intervention patients. The hospital-to-cannulation interval and procedural duration show that reaching an ECMO-capable hospital does not immediately restore perfusion. These delays plausibly reduced the opportunity to preserve neurological function, but they were not randomised exposures. INCEPTION cannot quantify the benefit that would have occurred with a shorter interval, nor establish a safe or effective universal time threshold.¹

The multicentre design also creates a tension between relevance and consistency. Participating hospitals had experience with venoarterial ECMO, but cannulation teams, locations and recruitment volumes varied. Some centres joined late, and pandemic interruptions disrupted enrolment. Experience with planned or urgent circulatory support does not ensure frequent practice in cannulation during ongoing cardiac arrest. These features help explain why results from an established specialist programme may not transfer automatically to a wider network. They do not, without a credible interaction analysis, identify which centres or operators produced a treatment benefit.^{1,3}

Eligibility was selective, and screening information was incomplete across the network. In the five centres providing screening data, 2,107 patients were assessed and 113 were

randomised. The screening figure also identifies 44 potential candidates who were not included because of logistical problems or failure to recognise eligibility. Their small proportion of all screened arrests should not be mistaken for a negligible proportion of eligible patients. Alongside the upper age limit, predominantly male population and near-universal early CPR, this limits extrapolation to older patients, prolonged unwitnessed arrests, initial non-shockable rhythms or systems with substantially different prehospital performance.³

Outcome ascertainment was stronger than reliance on unblinded treating clinicians, but a nonstandardised neurological inquiry and a coarse CPC dichotomy cannot describe all disability experienced by survivors. Treatment withdrawal remained dependent on local clinical practice. In the ECPR group, more patients reached intensive care and could subsequently have neurological injury recognised; many controls died during unsuccessful resuscitation. The differing recorded reasons for withdrawal therefore cannot establish that ECPR caused more brain injury. Contemporary correspondence raised the relevance of post-resuscitation care and the difficulty of reproducing specialist-centre results, but these remain explanations to test rather than demonstrated mechanisms of the primary outcome.^{1,5}

Safety needs the same attention to observation time and competing death. More intervention patients survived long enough to undergo invasive procedures, receive prolonged support and accumulate complications. Conversely, early control deaths prevented many later adverse events from occurring or being observed. Neither the mean event count nor a statement that event rates were similar can establish safety equivalence. Major bleeding and limb ischaemia are clinically important burdens, while the revised protocol's decision not to register every expected critical illness event limits the completeness of a simple adverse-event comparison.^{2,3}

Finally, the documentation supports a reliable appraisal of the primary result but leaves some reporting uncertainty. The final protocol, amendment summary and principal report differ in the target number and chronology of sample-size expansion. The supplied statistical analysis plan (SAP) has no visible approval date and does not reproduce every analytical detail in the publication, including the final alpha adjustment. Several prespecified metabolic, rehabilitation and temperature outcomes were unavailable or deferred, with reasons given in the supplement. These limitations warrant transparency rather than an allegation of selective analysis. The defensible conclusion remains that this multicentre implementation did not establish improved neurological survival, while leaving a clinically important treatment effect unresolved.^{1,2,3,4}

Body of Evidence

The early clinical rationale was supported principally by non-randomised comparisons. In the Japanese SAVE-J study published in 2014, Sakamoto and colleagues compared a treatment bundle including ECPR, hypothermia and intra-aortic balloon support at 26 hospitals with conventional care at 20 hospitals for shockable out-of-hospital arrest. Favourable neurological outcome occurred in 32 of 260 patients (12.3%) versus three of 194 (1.5%) at one month, and 29 of 260 (11.2%) versus five of 194 (2.6%) at six months. These were large associations, but treatment strategy was linked to hospital and care bundle. Differences in selection, institutional capability and accompanying treatment could contribute to the apparent benefit. The study justified randomised testing without establishing the causal effect of ECMO alone.⁶

The Minnesota ARREST trial, published in 2020, supplied direct randomised evidence in a tightly organised single-centre system. This is a different study from the London hospital-destination ARREST trial. Adults aged 18–75 years with refractory ventricular fibrillation after three shocks, mechanical CPR and an estimated transport time under 30 minutes were randomised at hospital arrival. Thirty patients were assigned equally; one intervention patient withdrew. Survival to discharge was six of 14 with ECMO-facilitated resuscitation versus one of 15 with standard advanced life support. The estimated absolute difference was 36.2 percentage points, with a 95% Bayesian credible interval of 3.7–59.2 and a posterior probability of superiority of 0.9861. The trial stopped at its first planned interim analysis. Its large effect was encouraging, but the small sample, early stopping and specialised delivery system limited the precision and transferability of the estimate.⁷

Prague OHCA, published in 2022, evaluated an invasive resuscitation strategy in a broader rhythm population. Patients were randomised during ongoing resuscitation to intra-arrest transport, possible ECPR and immediate invasive assessment, or continued standard resuscitation on scene. After withdrawals, 256 patients were analysed. At 180 days, favourable neurological survival occurred in 39 of 124 patients (31.5%) assigned to the invasive strategy and 29 of 132 (22.0%) assigned to standard care: odds ratio 1.63, 95% confidence interval 0.93–2.85; absolute difference 9.5 percentage points, 95% confidence interval –1.3 to 20.1; $P=0.09$. Recruitment stopped under a prespecified futility rule. The estimate remained compatible with worthwhile benefit. Unlike INCEPTION, however, the assigned strategy also changed the initial decision to transport, and the population included non-shockable arrests. The comparison therefore informs the whole resuscitation pathway rather than providing an interchangeable estimate of ECPR efficacy.⁸

INCEPTION subsequently tested an ECPR option across a larger network in which both groups could undergo intra-arrest transport. Its smaller observed separation in neurological survival should be interpreted alongside these differences in allocation timing, population and delivery. The three trials do not form a simple sequence from efficacy to refutation. Each asks what happens when a particular service offers an invasive rescue strategy at a particular stage of refractory arrest. Control-group recovery, the proportion requiring extracorporeal support and the speed with which that support is established all influence the resulting comparison.^{1,7,8}

A prespecified INCEPTION per-protocol analysis became available online on 3 November 2023. It excluded additional patients, including those with delayed cannulation, spontaneous circulation before hospital arrival and crossover treatment, leaving 33 intervention patients and 48 controls. Favourable neurological survival at 30 days was 15% versus 9%, with an adjusted odds ratio of 1.9 (95% confidence interval 0.4–9.3; $P=0.393$). A supplementary Bayesian analysis estimated an 84% probability of any benefit. These results remained highly uncertain. More fundamentally, selecting patients by adherence and events after allocation no longer preserves the original randomised comparison. The analysis can describe outcomes under closer protocol adherence but cannot prove that eliminating delays would reproduce its point estimate in all eligible patients.⁹

By the end of 2023, the evidence supported continued uncertainty about which patients and systems gain sufficient benefit to justify ECPR's burdens. Observational promise and the Minnesota signal remained relevant, but broader randomised implementation had not established consistent improvement in neurological survival. INCEPTION's practical contribution is to make delivery part of the appraisal: selecting patients, preserving CPR quality, preparing an experienced team, achieving flow and providing subsequent critical care are all components of the intervention. Its findings do not justify assuming that access to an ECMO circuit alone will improve outcomes.¹

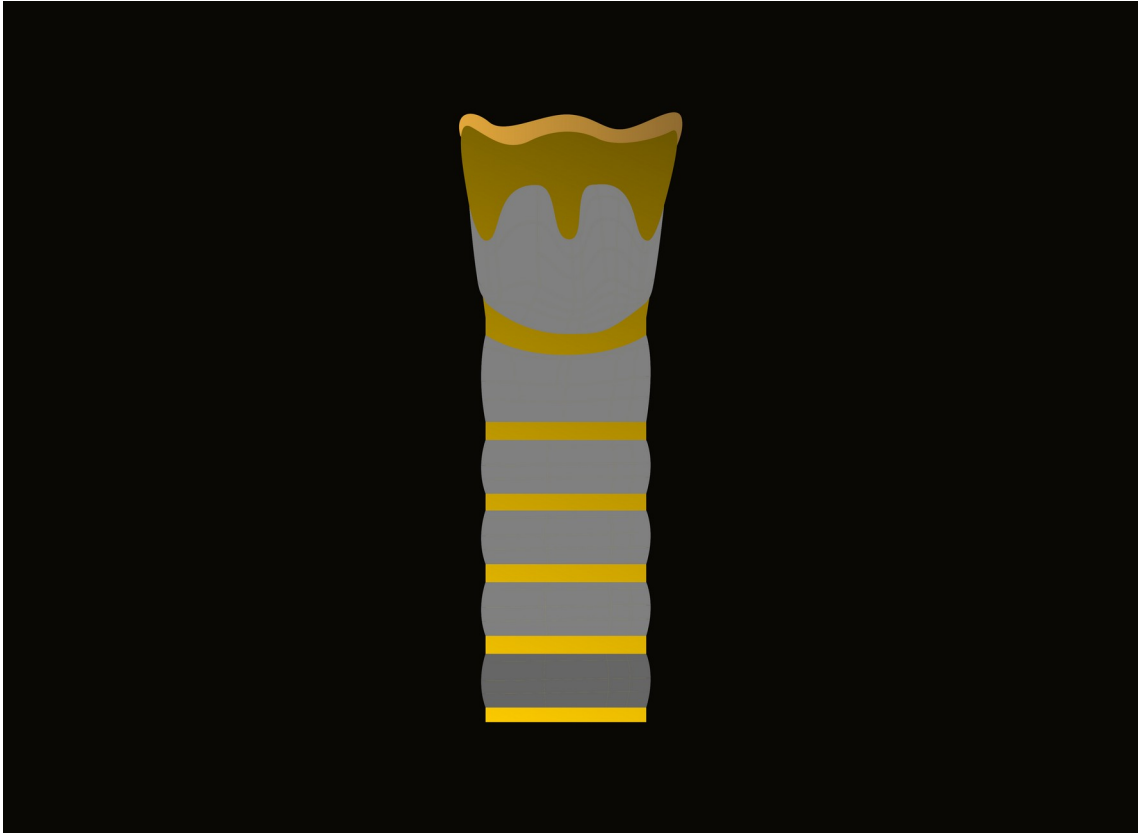
Does INCEPTION support routine ECPR for refractory shockable out-of-hospital cardiac arrest?

No. It did not establish improved neurological survival; benefit in selected patients within experienced, rapid-response systems remains uncertain.

References

1. Suverein MM, Delnoij TSR, Lorusso R, Brandon Bravo Bruinsma GJ, Otterspoor L, Elzo Kraemer CV, et al. Early Extracorporeal CPR for Refractory Out-of-Hospital Cardiac Arrest. *N Engl J Med*. 2023;388(4):299–309.
2. INCEPTION Trial Investigators. Trial protocol: original version 3, 13 January 2017; final version 3.9, 15 December 2020; and summary of amendments. Supplementary material to: Early Extracorporeal CPR for Refractory Out-of-Hospital Cardiac Arrest. *N Engl J Med*. 2023;388:299–309.
3. Suverein MM, Delnoij TSR, Lorusso R, Brandon Bravo Bruinsma GJ, Otterspoor L, Elzo Kraemer CV, et al. Supplementary appendix to: Early Extracorporeal CPR for Refractory Out-of-Hospital Cardiac Arrest. *N Engl J Med*. 2023;388:299–309.
4. INCEPTION Trial Investigators. Statistical analysis plan for the INCEPTION trial, based on protocol version 3.9, and summary of changes. Supplementary material to: Early Extracorporeal CPR for Refractory Out-of-Hospital Cardiac Arrest. *N Engl J Med*. 2023;388:299–309.
5. Cho SM, Geocadin RG, Whitman GJ. Extracorporeal CPR for Out-of-Hospital Cardiac Arrest. *N Engl J Med*. 2023;388(20):1915–1916.
6. Sakamoto T, Morimura N, Nagao K, Asai Y, Yokota H, Nara S, et al. Extracorporeal cardiopulmonary resuscitation versus conventional cardiopulmonary resuscitation in adults with out-of-hospital cardiac arrest: a prospective observational study. *Resuscitation*. 2014;85(6):762–768.
7. Yannopoulos D, Bartos J, Raveendran G, Walser E, Connett J, Murray TA, et al. Advanced reperfusion strategies for patients with out-of-hospital cardiac arrest and refractory ventricular fibrillation (ARREST): a phase 2, single centre, open-label, randomised controlled trial. *Lancet*. 2020;396(10265):1807–1816.
8. Belohlavek J, Smalcova J, Rob D, Franek O, Smid O, Pokorna M, et al. Effect of Intra-arrest Transport, Extracorporeal Cardiopulmonary Resuscitation, and Immediate Invasive Assessment and Treatment on Functional Neurologic Outcome in Refractory Out-of-Hospital Cardiac Arrest: A Randomized Clinical Trial. *JAMA*. 2022;327(8):737–747.
9. Ubben JFH, Suverein MM, Delnoij TSR, Heuts S, Winkens B, Gabrio A, et al. Early extracorporeal CPR for refractory out-of-hospital cardiac arrest – A pre-planned per-protocol analysis of the INCEPTION-trial. *Resuscitation*. 2024;194:110033. Published online 3 November 2023.

Possible Reporting Inconsistencies. The reported mean times to return of spontaneous circulation, 43 ± 20 and 49 ± 19 minutes, are assigned to opposite treatment groups in the Results prose and Table 2. This leaves the direction of the between-group timing difference unclear.¹



Airway Trials

NICO

Freund Y, Viglino D, Cachanado M, Cassard C, Montassier E, Douay B, et al. Effect of noninvasive airway management of comatose patients with acute poisoning: a randomized clinical trial. JAMA. 2023;330(23):2267–2274.

Introduction

A patient who is deeply unconscious after poisoning presents an immediate airway decision. Tracheal intubation can secure ventilation and facilitate treatment when airway protection is inadequate. It also exposes the patient to induction drugs, neuromuscular blockade, procedural complications and subsequent mechanical ventilation. The difficulty is identifying patients who require these interventions while avoiding a sequence of intensive care treatment in those whose consciousness will recover with supportive care. A low Glasgow Coma Scale (GCS) score describes the depth of impaired consciousness, but does not by itself establish the balance of these competing risks.¹

The familiar threshold of a GCS score of eight or less has often been extended from traumatic brain injury to poisoning. The underlying situations can differ substantially. Poisoning may produce profound but reversible depression of consciousness, while the presence of respiratory failure, seizures, shock or an associated cerebral injury changes both urgency and prognosis. Earlier observational work suggested that selected poisoned patients could be observed without intubation, but clinician selection made the apparent safety of this approach difficult to interpret causally. A service choosing observation for its most reassuring patients cannot establish what would have happened if equivalent patients had instead been intubated.¹

NICO addressed this uncertainty by randomising an airway-management strategy. Its intervention was active surveillance with defined indications for rescue intubation. The comparison was usual clinical discretion, under which intubation remained optional. This distinction matters: the trial tested whether a structured policy could reduce unnecessary invasive treatment in selected patients, rather than comparing intubation with its complete prohibition. Its hierarchical outcome also combined survival with time spent in intensive care and hospital, requiring careful separation of clinical recovery from the consequences of the initial airway decision.¹

Synopsis

NICO was a pragmatic, open-label, individually randomised trial conducted at 20 emergency departments and one intensive care unit (ICU) in France. Sixteen emergency departments also enrolled patients through physician-led prehospital emergency teams. Recruitment ran from 16 May 2021 to 12 April 2023, with final follow-up on 12 May. The French Ministry of Health funded the study and Assistance Publique–Hôpitaux de Paris sponsored it. An independent data and safety monitoring board oversaw safety; no interim efficacy analysis was planned. Consent could be obtained after emergency enrolment when the patient recovered sufficiently, or from an appropriate representative.^{1,2}

Adults were eligible if an emergency physician suspected acute poisoning by alcohol, recreational drugs or medication and assessed a GCS score below nine. Patients requiring immediate intubation were excluded, including those with respiratory distress, shock, seizures or suspected brain injury. Suspected poisoning with specified cardiotoxic drugs, including beta blockers, calcium channel blockers and angiotensin-converting enzyme inhibitors, also excluded participation. The protocol additionally excluded QRS or QT prolongation and isolated poisoning for which an antidote was available. Pregnancy, breastfeeding and several legal or administrative circumstances were further exclusions. Toxicological confirmation was not required for enrolment.^{1,2}

Patients were assigned equally using centre-stratified, balanced-block randomisation and sealed envelopes distributed to participating clinical teams. Block size was undisclosed. In the restricted-intubation group, intubation was withheld unless an emergency criterion developed: seizure, vomiting, respiratory distress with oxygen saturation below 90% persisting despite nasal oxygen, or systolic blood pressure below 90 mmHg persisting after one litre of crystalloid. The assigned strategy continued for four hours or until the GCS exceeded eight, whichever occurred first. Thereafter, management returned to clinical discretion.¹

Observation required a monitored resuscitation environment with continuous physical presence of a nurse or physician. The protocol specified repeated assessment of vital signs and GCS every 30 minutes. In the control group, the treating physician decided whether and when to intubate. When intubation was needed, rapid sequence induction, preoxygenation and confirmation using capnography were recommended; the laryngoscope and airway adjunct were chosen clinically. The study did not randomise an induction drug, drug dose or laryngoscopy device. No patient underwent gastric evacuation or received activated charcoal before intubation.^{1,2}

The primary outcome ranked in-hospital death, ICU length of stay and hospital length of stay, in that order, with observation truncated at 28 days. Each intervention patient was compared with each control patient. Survival took precedence; among survivors, shorter ICU stay determined the winner, followed by hospital stay when ICU stays were equal. Transfer to a psychiatric ward counted as hospital discharge for this assessment. The protocol specified a Finkelstein–Schoenfeld analysis and the later statistical analysis plan (SAP) described complementary win-ratio and win-odds estimates, including centre-stratified analyses. The publication presented the primary effect as a win ratio with a 95% confidence interval.^{1,2,3}

Simulation suggested that 100 patients per group would provide 98% power at a two-sided 5% significance level. Assumptions included 3% mortality in each group, mean ICU stays of zero versus one day and mean hospital stays of two versus four days. The recruitment target was 240 after allowing for missing follow-up and refusal of continued participation. The primary analysis used a modified intention-to-treat population. Secondary outcomes included the individual components, ICU admission, mechanical ventilation, pneumonia within 48 hours and intubation-related adverse events. Secondary comparisons used models accounting for centre; their confidence intervals were not adjusted for multiple testing.^{1,2,3}

Of 237 randomised patients, 119 were assigned to restricted intubation and 118 to usual care. Three and nine, respectively, were excluded because of refusal to use their data or legal protection, leaving 116 and 109 in the primary analysis. The total screened population was not recorded. Median age was 33 and 34 years, and median GCS was six in both groups. Women comprised 85 of 225 patients. Alcohol was implicated in approximately two-thirds; benzodiazepine exposure was also common, and multiple substances could be involved.¹

The hierarchical primary outcome favoured restricted intubation: win ratio 1.85, 95% confidence interval 1.33–2.58; $P < 0.001$. There were no deaths in either group. Across 12,644 patient pairs, ICU stay produced 7,015 wins and 2,998 losses for the intervention; hospital stay resolved most remaining ties, giving 8,166 wins, 4,404 losses and 74 final ties. Median ICU stay was zero hours versus 24 hours, with a modelled rate ratio of 0.39 (95% confidence interval 0.24–0.66). Median hospital stay was 21.5 versus 37 hours; its rate ratio was 0.74 (95% confidence interval 0.53–1.03). The centre-stratified win ratio was similar at 1.83 (95% confidence interval 1.29–2.60).^{1,4}

During initial care, 19 of 116 intervention patients (16.4%) and 63 of 109 controls (57.8%) were intubated. Sixteen intervention patients met an emergency criterion; the other three represented departures from the restricted strategy. ICU admission occurred in 46

of 116 (39.7%) versus 72 of 109 (66.1%), with a reported model-based difference of -29.2 percentage points (95% confidence interval -41.0 to -17.4). Mechanical ventilation, including noninvasive ventilation, was recorded in 21 of 116 versus 65 of 109. Its median duration was zero versus six hours. Four additional patients were intubated later during their hospital course.^{1,4}

Pneumonia occurred in eight of 116 intervention patients and 16 of 109 controls: odds ratio 0.43, 95% confidence interval 0.18–1.05. Intubation-related adverse events were recorded in seven of 113 patients with available data versus 16 of 107. The reported modelled absolute difference was -8.6 percentage points (95% confidence interval -16.6 to -0.7). No intubation-related cardiac arrest occurred. The per-protocol win ratio was 1.91 (95% confidence interval 1.37–2.68). In prespecified subgroup analyses, the win ratio was 1.70 (95% confidence interval 1.10–2.64) for GCS below seven and 1.42 (95% confidence interval 0.88–2.32) for the subgroup described as alcohol, benzodiazepine, gamma-hydroxybutyrate or gamma-butyrolactone poisoning.^{1,4}

Critique

The most consequential feature of NICO is that its positive primary result was driven by ICU use and duration, with no contribution from mortality. In the French pathway, patients intubated before hospital arrival were taken directly to intensive care. Avoiding intubation could therefore prevent ICU admission by changing the care pathway, even without altering the underlying toxicological illness. That is a clinically useful effect when invasive treatment is unnecessary, but the hierarchical label should not obscure what improved. NICO demonstrated a favourable overall ranking of hospital outcomes, principally through less intensive care exposure; it did not demonstrate improved survival.¹

The win ratio also needs its proper scale. A value of 1.85 means that intervention patients won more cross-group comparisons than they lost under the specified hierarchy. It does not mean that 85% of patients benefited, that mortality fell by 85%, or that every patient left hospital sooner. Most comparisons were decided by ICU stay. Among pairs still tied at that stage, hospital stay generated more losses than wins for the intervention. That conditional comparison is not a separate randomised subgroup analysis, but it illustrates why the overall result cannot be translated into uniform improvement in every component. The confidence interval for the separate hospital-stay estimate included no effect.¹

Randomisation provides an important advance over observational comparisons between intubated and non-intubated patients. The substantial separation in actual intubation

rates shows that assignment changed treatment. However, the control strategy remained discretionary: more than two-fifths of controls avoided initial intubation. The result therefore estimates the value of a structured restriction relative to the participating clinicians' usual threshold. A service already practising similar selective observation may obtain a smaller additional reduction. Conversely, implementation in a service that intubates almost every poisoned patient below the GCS threshold could change treatment more extensively than NICO itself tested.¹

The restriction was inseparable from monitoring and rescue capability. Sixteen intervention patients required intubation after an emergency criterion arose, including patients deteriorating two to four hours after enrolment. This was not passive observation in an unmonitored bed. Continuous attendance, repeated clinical assessment and immediate access to airway expertise were part of the intervention's conditions. Reducing ICU admission can transfer work to emergency teams rather than abolishing it. Staffing demands and the ability to sustain surveillance are particularly relevant when applying these findings to crowded emergency departments or prehospital systems without an attending physician.^{1,2,5}

Open treatment assignment was unavoidable, but it affects interpretation of resource outcomes. ICU admission, readiness for discharge, sedation and subsequent care were decided by clinicians who knew the allocated strategy. Some differences may reflect the intended removal of an unnecessary intervention, while others may reflect altered clinical thresholds. The trial cannot separate those mechanisms. Pneumonia assessment used a specified clinical and radiological definition, which improves consistency, but the protocol placed adjudication with local investigators rather than an independent blinded panel. This leaves scope for differential investigation and classification in a small, unblinded trial.^{1,2}

The safety results are reassuring within their limits. Fewer assigned patients experienced an intubation-related event when fewer patients underwent intubation. That is a valid strategy-level benefit, but it does not show that rescue intubation itself was safer than elective intubation. Comparisons restricted to patients who were eventually intubated lose randomisation because deterioration and clinical selection determine entry into that group. The same problem affects the very large exploratory win ratio favouring patients who were never intubated. It largely contrasts patients following different clinical courses and cannot estimate the causal effect of withholding intubation.^{1,3}

Nor do zero deaths establish equivalent survival. The sample-size calculation targeted the hierarchical outcome and expected its discrimination to arise from differences in length of stay. It did not provide adequate information about uncommon fatal

complications. The pneumonia estimate favoured restriction, but its confidence interval included no difference. A significant pneumonia result in the per-protocol analysis cannot displace the more inclusive randomised comparison, particularly after exclusion of protocol departures and without adjustment for the many secondary analyses. The absence of a demonstrated increase in pneumonia should therefore support cautious reassurance rather than a claim that aspiration risk has been eliminated.^{1,4}

Selection constrains the clinical boundary. These were predominantly young adults with alcohol or sedative involvement, assessed as not already requiring immediate intubation. Respiratory failure, haemodynamic instability, seizures, suspected cerebral injury and important cardiotoxic exposures were exclusions. The trial does not establish the strategy for patients with those features, for children, or for all forms of nontraumatic coma. Suspected poisoning also remains a clinical diagnosis: a low GCS after alcohol exposure does not exclude an additional cerebral injury. Incomplete screening data prevent a reliable estimate of what proportion of all comatose poisoned patients met the study's criteria.^{1,2}

The subgroup findings do not supply a new GCS threshold or identify a particular intoxicant that benefits most. Statistical significance in the lower-GCS subgroup and its absence in the restricted-drug subgroup are not evidence of different treatment effects. Those questions require a credible interaction assessment and substantially more information than these small subgroups provide. The clinically relevant distinction remains the presence or absence of an immediate airway indication, together with the capacity to observe and rescue, rather than a numerical consciousness threshold considered in isolation.^{1,3}

Finally, several documentation issues deserve explicit attention. The supplied SAP is dated July 2023, after recruitment and follow-up, while stating that it preceded database lock. The protocol and SAP place the Finkelstein–Schoenfeld method before complementary win-ratio analyses, whereas the publication emphasises the win ratio. The available documents do not establish when investigators first had access to unblinded outcome comparisons. Post-randomisation exclusions were also unequal between groups, and the peri-intubation hypotension definition differs between the Methods and Table 2, as noted below. These issues warrant clarification without invalidating the clear treatment separation or the resource-use result. They reinforce the need to avoid claims that depend on fragile secondary estimates or unverified analytical chronology.^{1,2,3}

Body of Evidence

The observational evidence preceding NICO was more effective at challenging an automatic GCS threshold than at determining the optimal alternative. Duncan and Thakore's 2009 prospective study followed 73 patients admitted to a Scottish emergency department short-stay ward with poisoning-related impairment of consciousness. Twelve had a GCS of eight or less; none of these patients developed clinically significant aspiration or required intubation. However, patients already intubated in the resuscitation area and transferred to intensive care were excluded. The reassuring experience therefore concerned a clinically selected ward population, supported by experienced emergency staff and ready access to resuscitation equipment. It demonstrated feasibility, with limited precision and no randomised estimate of benefit or harm.⁶

A separate 2009 prospective comparison by Donald and colleagues included 12 intubated and 14 conservatively managed poisoned patients in a Scottish emergency department. The non-intubated group had a GCS of eight or less, while allocation reflected clinical judgement. The investigators concluded that experienced assessment of airway protection and ventilation was more informative than GCS alone. The study helped formulate the question that NICO later addressed, but its small size and treatment selection prevented a causal comparison. It also illustrates a methodological difficulty: predicting which patients clinicians intubate is not the same as identifying which patients would benefit from intubation.⁷

The 2020 report by van Helmond and Gresnigt examined 209 presentations with suspected gamma-hydroxybutyrate or gamma-butyrolactone intoxication at a Dutch emergency department. Major adverse events occurred in five patients (2.4%, 95% confidence interval 0.8–5.5), while 1.4% required intubation and subsequent ICU admission. There were no deaths or recorded signs of aspiration pneumonia. Minor airway obstruction and hypoxia were appreciably more common and were managed with conservative measures. This retrospective experience supported active supportive care for a particular toxicological presentation; it did not establish comparative safety for all poisons or remove the need for intervention when basic measures failed.⁸

Orso and colleagues' 2020 systematic review brought together 13 observational studies spanning nontraumatic, traumatic and mixed populations. Results concerning aspiration were inconsistent, and heterogeneity prevented a quantitative meta-analysis. The review found insufficient evidence to establish whether intubation reduced aspiration in patients with impaired consciousness. Its contribution was to expose the evidential weakness behind a widely used threshold, rather than to prove that intubation was harmful or unnecessary. The differing causes of coma, opportunities for aspiration

before treatment and clinical reasons for selecting intubation made a randomised strategy comparison especially valuable.⁹

NICO supplied that comparison in a defined poisoning population. Its contribution was not simply another report of patients surviving without a tracheal tube. It showed that assigning a monitored, restrictive policy materially changed airway treatment and reduced intensive care exposure compared with contemporaneous usual care. This addresses confounding by initial treatment selection more directly than the preceding cohorts. It leaves the limits imposed by its setting, exclusions, sample size and open resource decisions, and it does not resolve every concern about aspiration or uncommon catastrophic deterioration.¹

The accompanying editorial emphasised the potential value of avoiding unnecessary invasive care, while recognising the practical importance of patient selection and monitoring. At the end of 2023, the defensible clinical change was to question intubation based solely on GCS in an otherwise eligible poisoned adult and to replace that reflex with structured assessment and observation. The evidence did not justify a universal non-intubation rule. The remaining research need was to establish how consistently this approach could be delivered across different emergency systems, how much clinical surveillance it required and whether uncommon serious harms remained acceptably infrequent.^{1,5}

Should a GCS below nine alone prompt intubation after acute poisoning?

NICO supports monitored observation with explicit rescue criteria in selected adults without an immediate indication for intubation. It reduced intubation and ICU exposure, but did not establish a survival benefit or safety across all causes of coma.

References

1. Freund Y, Viglino D, Cachanado M, Cassard C, Montassier E, Douay B, et al. Effect of noninvasive airway management of comatose patients with acute poisoning: a randomized clinical trial. *JAMA*. 2023;330(23):2267–2274.
2. NICO investigators. Non-invasive airway management of comatose poisoned emergency patients: study protocol, version 5.0, 28 December 2022. Supplement 1 to Freund et al. *JAMA*. 2023;330:2267–2274.
3. NICO investigators. Statistical analysis plan, version 1, 4 July 2023. Supplement 2 to Freund et al. *JAMA*. 2023;330:2267–2274.

4. Freund Y, Viglino D, Cachanado M, Cassard C, Montassier E, Douay B, et al. Supplementary online content to: Effect of noninvasive airway management of comatose patients with acute poisoning: a randomized clinical trial. *JAMA*. 2023;330:2267–2274.
5. Qasim Z, Perrone J, Delgado MK. The value of not intubating comatose patients with acute poisoning. *JAMA*. 2023;330(23):2253–2254.
6. Duncan R, Thakore S. Decreased Glasgow Coma Scale score does not mandate endotracheal intubation in the emergency department. *J Emerg Med*. 2009;37(4):451–455.
7. Donald C, Duncan R, Thakore S. Predictors of the need for rapid sequence intubation in the poisoned patient with reduced Glasgow coma score. *Emerg Med J*. 2009;26(7):510–512.
8. van Helmond LPFM, Gresnigt FMJ. Safety of withholding intubation in gamma-hydroxybutyrate- and gamma-butyrolactone-intoxicated coma patients in the emergency department. *Eur J Emerg Med*. 2020;27(3):223–227.
9. Orso D, Vetrugno L, Federici N, D’Andrea N, Bove T. Endotracheal intubation to reduce aspiration events in acutely comatose patients: a systematic review. *Scand J Trauma Resusc Emerg Med*. 2020;28:116.

Possible Reporting Inconsistencies. The abstract appears to omit a minus sign before the 8.6-percentage-point adverse-event difference: its confidence interval is –16.6 to –0.7, and Table 2 gives –8.6. Peri-intubation hypotension is also defined as systolic pressure below 65 mmHg in the Methods but below 90 mmHg in Table 2. The supplied protocol and SAP do not resolve that adverse-event threshold discrepancy.^{1,2,3}

DEVICE

Prekker ME, Driver BE, Trent SA, Resnick-Ault D, Seitz KP, Russell DW, et al.; DEVICE Investigators and the Pragmatic Critical Care Research Group. Video versus Direct Laryngoscopy for Tracheal Intubation of Critically Ill Adults. *N Engl J Med*. 2023;389(5):418–429.

Introduction

Tracheal intubation in critical illness requires more than placing a tube in the correct anatomical position. Hypoxaemia and circulatory instability may develop while clinicians secure the airway, and improving the mechanics of intubation need not prevent every physiological complication. In the international INTUBE observational study, 45.2% of 2,964 patients experienced a major adverse event within 30 minutes of starting intubation. Cardiovascular instability predominated. This burden makes the choice of the first intubating device consequential, while also requiring a distinction between procedural success and patient safety.¹

Direct laryngoscopy requires a line of sight from the operator to the glottis. A video laryngoscope places a camera near the blade tip and displays the larynx on a screen. This can improve visualisation without requiring the same alignment of the airway. The endotracheal tube must nevertheless negotiate a separate path into the trachea; an excellent image does not guarantee easy tube delivery. Blade geometry, the use of a stylet or bougie, and familiarity with the device may therefore influence the practical benefit of the camera.²

Before DEVICE, randomised trials in critically ill adults had not consistently shown better first-attempt success with video laryngoscopy. FELLOW and MACMAN demonstrated the distinction between an improved view and successful intubation, and MACMAN raised concern about severe complications. These findings left an important uncertainty despite the intuitive appeal of video assistance and its availability in emergency airway practice.^{3,4}

DEVICE asked whether starting with a video laryngoscope, within contemporary emergency-department and intensive-care practice, would improve successful tracheal intubation on the first attempt. The hypothesis concerned the initial procedural strategy, with clinicians retaining responsibility for subsequent attempts and rescue. Whether any improvement would also reduce severe peri-intubation complications was a separate, clinically important question.⁵

Synopsis

DEVICE was a pragmatic, multicentre, unblinded, parallel-group superiority trial conducted in seven emergency departments and ten intensive care units (ICUs) at 11 US medical centres. Recruitment ran from 19 March to 17 November 2022. The US Department of Defense funded the trial, with additional research support acknowledged; the published protocol states that the funding sources had no role in trial design, conduct, analysis or reporting. Enrolment proceeded under a waiver of informed consent.^{2,5}

Eligible patients were adults undergoing planned orotracheal intubation with a laryngoscope in a participating unit, performed by a clinician expected routinely to intubate there. Known pregnancy, imprisonment and age below 18 years were exclusions. Patients were also excluded when intubation was too urgent to permit trial procedures, or the operator considered either laryngoscope required or contraindicated. Computer-generated allocation used variable-sized permuted blocks, stratified by site, with a 1:1 ratio and sequentially numbered opaque envelopes. Operators and procedural observers were not blinded.^{2,6}

The video group was assigned a device with a camera and screen for the first laryngoscopy attempt. Brand and blade geometry were chosen by the operator; screen viewing was encouraged. The direct group was assigned a conventional laryngoscope without a camera or screen. Device crossover on the first attempt was permitted for patient safety. After an unsuccessful attempt, clinicians could choose any rescue approach. Positioning, preoxygenation, induction drugs, neuromuscular blockade, and selection of a bougie or stylet remained discretionary. Thus, randomisation specified the initial laryngoscope rather than a complete airway-management bundle.^{2,6}

The primary outcome was successful intubation using one insertion of the laryngoscope blade and either one insertion of the endotracheal tube or one bougie insertion followed by one tube insertion. A trained independent observer recorded these events, with operator reporting substituted when observer data were unavailable. Discordance between observer and operator classified the attempt as unsuccessful. Tracheal placement was confirmed according to local protocols including detection of exhaled carbon dioxide.⁶

The single secondary outcome was a composite of severe complications from induction until two minutes after intubation: oxygen saturation below 80%, systolic pressure below 65 mm Hg, new or increased vasopressor administration, or cardiac arrest. Arrests were classified according to whether death followed within one hour after intubation. Exploratory outcomes included procedural duration, laryngeal view, aspiration,

oesophageal intubation, dental injury and clinical outcomes through day 28. ICU-free and ventilator-free days assigned zero to patients who died; follow-up was censored at hospital discharge or day 28.⁶

The planned sample was 2,000, allowing 4% missing data. An evaluable sample of 1,920 provided 90% power to detect a five-percentage-point improvement from an assumed 80% direct-laryngoscopy success rate, at a two-sided 0.05 significance level, with allowance for correlation within sites. One interim analysis after 1,000 patients used a Haybittle–Peto efficacy boundary of $P \leq 0.001$. The principal comparison used a chi-square test by assigned group. Prespecified sensitivity analyses addressed covariate adjustment and site effects, crossover, observer-only ascertainment and operators with more balanced prior exposure to both devices.^{2,6}

Of 1,947 patients assessed, 1,420 were randomised. Important exclusions included 225 requiring intubation too urgently and 82 for whom a particular device was required. Three patients subsequently identified as prisoners had their data removed, leaving 705 in the video group and 712 in the direct group. The interim analysis showed success in 425/494 versus 365/506 patients and crossed the stopping boundary. Recruitment continued while follow-up and review were completed, adding 420 randomised patients before the monitoring board recommended stopping on 17 November.⁶

The median age was 55 years; 69.7% of intubations occurred in emergency departments. Altered mental status and respiratory failure accounted for 45.3% and 30.4% of indications, respectively, and 23.9% of patients had traumatic injury. There were 387 operators, contributing a median of two trial intubations each. Emergency medicine residents or critical care fellows performed 91.5% of procedures. The median prior intubation experience was 50 procedures, and the median proportion previously performed with video laryngoscopy was 0.69.⁵

All 705 patients assigned video laryngoscopy received it initially; 704/712 assigned direct laryngoscopy received their allocated device, with eight receiving video laryngoscopy. Standard-geometry video blades were used in 607/705 video-group procedures and hyperangulated blades in 98/705. Screen viewing was documented in 674/690 evaluable video-group procedures. Almost all patients received preoxygenation, and sedation and neuromuscular blockade were used in approximately 96% of evaluable cases. A stylet or bougie was used whenever tube placement was attempted during the first laryngoscopy; some initial laryngoscopies ended without tube insertion.^{5,6}

First-attempt success occurred in 600/705 patients (85.1%) with video versus 504/712 (70.8%) with direct laryngoscopy: an absolute improvement of 14.3 percentage points

(95% confidence interval [CI] 9.9–18.7; $P < 0.001$). The adjusted odds ratio was 2.55 (95% CI 1.94–3.30). Observer-only and crossover sensitivity analyses were consistent. Among 827 procedures performed by operators whose previous video use represented 25–75% of intubations, the absolute improvement was 13.5 percentage points (95% CI 7.7–19.4). All analysed patients had a primary outcome; operator reporting alone supplied eight outcomes.^{5,6}

Severe complications occurred in 151/705 (21.4%) versus 149/712 (20.9%), an absolute difference of 0.5 percentage points (95% CI –3.9 to 4.9). Oxygen saturation below 80% occurred in 64/658 (9.7%) versus 69/659 (10.5%); systolic pressure below 65 mm Hg in 20/624 (3.2%) versus 29/644 (4.5%); and new or increased vasopressor use in 91/705 (12.9%) versus 87/712 (12.2%). Each group had three peri-intubation cardiac arrests. Missing measurements for individual components were treated as absence of that complication when constructing the composite.^{5,6}

Exploratory findings included a grade 1 laryngeal view in 76.3% versus 44.7% and a median interval from initial blade insertion to successful intubation of 38 versus 46 seconds, a difference of –8 seconds (95% CI –12 to –4). Oesophageal intubation occurred in six versus nine patients, dental injury in three versus two, and operator-reported aspiration in seven versus 12. In-hospital death by day 28 occurred in 184/705 (26.1%) versus 191/712 (26.8%), an absolute difference of –0.7 percentage points (95% CI –5.5 to 4.0). Secondary and exploratory confidence intervals were not adjusted for multiplicity.⁵

Critique

DEVICE establishes a substantial improvement in first-attempt intubation, but it does not establish a reduction in severe complications or mortality. This distinction is central to its interpretation. Avoiding an unsuccessful attempt is worthwhile: it limits repeated airway manipulation and the need to change the initial plan. Nevertheless, procedural success is an intermediate outcome, and the 14.3-percentage-point benefit cannot simply be translated into an equivalent improvement in patient safety. The trial directly measured severe complications and found closely similar rates, with uncertainty extending to clinically important differences in either direction.⁵

The safety result is less secure than the completeness of its composite denominator suggests. Peri-intubation oxygen saturation was unavailable in 100 patients and systolic pressure in 149. The supplement treated missing component measurements as absence of that event, whereas the published analysis plan described complete-case analysis for outcomes. A patient with missing pressure could still meet the composite through vasopressor use or another component, but an unrecorded hypotensive episode could

be missed. This is not evidence that such episodes occurred; it is a reason not to equate a fully populated composite with fully observed physiology. The 95% CI, even before considering this assumption, permits almost five percentage points more severe complications with video laryngoscopy.^{2,6}

The procedural finding has stronger protection against missing-data bias. Only eight outcomes relied exclusively on operator reporting, and the observer-only analysis reproduced the primary absolute effect. The attempt definition was unusually strict, counting repeated tube or bougie insertion as failure even if the laryngoscope remained in place. This makes the absolute success rates sensitive to definition when comparing studies. However, the exploratory outcome based on a single blade insertion also favoured video laryngoscopy, 90.3% versus 77.3%. The conclusion therefore does not depend solely on counting multiple tube insertions as separate attempts.⁶

Blinding was impossible, and an independent observer was not a blinded adjudicator. Knowledge of assignment could influence operator persistence, assistance or the decision to abandon an attempt. Operator assessments of laryngeal view and reasons for failure are particularly vulnerable to expectation. Nonetheless, a prespecified, observable procedure count, very high adherence, site-stratified allocation and concordant sensitivity analyses make a large entirely artefactual primary effect less plausible. Patient and operator characteristics were reasonably balanced. The result is considerably more persuasive than a comparison based only on clinician preference or retrospective records.^{5,6}

Experience determines the population to which the average effect most readily applies. Most procedures were performed by trainees already familiar with video laryngoscopy; the median prior video proportion was about two-thirds. The question was not whether an unfamiliar video device outperformed a highly practised direct technique. Restricting analysis to more balanced prior device experience retained a sizeable benefit, weakening the explanation that familiarity alone produced the result. However, experience was self-reported, and this sensitivity analysis could not reproduce the skills of a consultant who has performed thousands of direct intubations.^{5,6}

The apparent benefit was larger with fewer than 25 previous intubations: 26.1 percentage points, compared with 5.9 percentage points among operators with more than 100. The latter interval ranged from -4.1 to 16.0 percentage points. These exploratory estimates support interest in training and experience but do not establish a threshold beyond which video laryngoscopy ceases to help. Nor does the imprecise estimate in more experienced operators prove equivalence. The supplementary experience plots count patient procedures, not an independent sample of operators at

each experience level, a distinction that matters when describing how much expert practice was represented.⁶

The trial tested a class of devices within an established system of tube delivery. Most video procedures used standard-geometry blades, and stylets or bougies accompanied attempted tube placement in both groups. DEVICE therefore provides its strongest evidence for this combination of available equipment and routine adjunct use. It cannot rank manufacturers, establish hyperangulated superiority, or justify assuming the same effect without tube-delivery aids. Failure attributed to an inadequate view fell markedly, while failure to pass a tube or bougie was similar. That pattern is consistent with improved visualisation helping overall success, but reasons for failure were operator-reported and sometimes missing; it is not a randomised analysis of mechanism.^{5,6}

Rescue use also defines the estimand. In the direct group, video laryngoscopy was documented on the final attempt in 123 patients whose first attempt had failed. Such rescue was permitted and clinically appropriate. It does not invalidate the first-attempt comparison, but it means subsequent clinical outcomes compare initial strategies with rescue available, rather than exclusive video versus exclusive direct laryngoscopy throughout. Convergence after an initial failure may reduce later outcome differences. Conversely, emergency-department predominance alone does not explain away the result: the ICU subgroup also showed a similar absolute improvement in first-attempt success.^{5,6}

Early stopping deserves qualification, rather than automatic dismissal. The boundary was prespecified and conservative; the first 1,000 patients already showed an absolute effect close to that in the final analysis. The additional enrolment during review provided considerable further information. These features support the direction of benefit, although stopping for an unusually favourable interim result can still overstate its magnitude. It also left fewer patients for uncommon safety outcomes. The protocol and statistical analysis plan were submitted during recruitment and available as a preprint before enrolment ended; peer-reviewed publication followed completion. The supplied history states that the statistical plan's content did not change between submission and publication.^{2,6}

Finally, clinician equipoise and the time required for enrolment selected the population. Patients requiring a particular device and those too unstable to permit trial procedures were excluded. A 72.9% enrolment proportion is reassuring within participating units, but it does not remove those important exceptions. With no demonstrated survival benefit, limited precision for rare airway injury and incomplete physiological

measurements, DEVICE supports choosing video laryngoscopy to improve first-attempt success in comparable practice. It leaves the management of severe anatomical difficulty, the most time-critical emergencies and highly experienced operators dependent on additional evidence and clinical judgement.^{5,6}

Body of Evidence

The earlier randomised literature was heterogeneous rather than uniformly unfavourable to video laryngoscopy. In the 150-patient FELLOW trial, conducted in a medical ICU with pulmonary and critical care fellows, video laryngoscopy improved the view but first-attempt success was 68.9% versus 65.8% with direct laryngoscopy. The adjusted odds ratio was 2.02 (95% CI 0.82–5.02). This wide interval did not exclude a worthwhile benefit, despite the non-significant result. DEVICE provides a much more precise estimate in a broader multicentre population, while retaining the distinction between visualisation, tube placement and subsequent complications.³

An emergency-department trial by Driver and colleagues randomised 198 patients in whom direct laryngoscopy had initially been planned, using the C-MAC device for either the direct or video approach. First-attempt success was 92% with video and 86% with direct laryngoscopy ($P=0.18$). Its smaller sample and selected clinical plan limited the certainty of the result. The higher direct-arm success also illustrates why a device effect cannot be transported solely by applying DEVICE's absolute difference to another unit. Starting success rates, operator skills and the precise comparator determine the available scope for improvement.⁷

MACMAN supplied the main multicentre ICU comparison before DEVICE. Across seven French ICUs, 371 patients were randomised to McGrath MAC video or Macintosh direct laryngoscopy. First-pass success was 67.7% versus 70.3%, an absolute difference of –2.5 percentage points (95% CI –11.9 to 6.9). A post hoc analysis of severe life-threatening complications favoured direct laryngoscopy, with rates of 9.5% versus 2.8%. This was a legitimate safety signal, although less secure than a prespecified primary finding. The protocol did not use a stylet for the first attempt, and failed tube passage despite a better view was an important problem. DEVICE's routine use of a stylet or bougie offers a plausible explanation for different results, but the two trials did not randomise that difference between them.⁴

The importance of tube delivery has separate randomised support. STYLETO included 999 critically ill adults in its modified intention-to-treat analysis across 32 ICUs. First-attempt success was 78.2% with an endotracheal tube and stylet versus 71.5% with a tube alone, an absolute difference of 6.7 percentage points (95% CI 1.4–12.1).

Complication rates were not significantly different. This supports treating the adjunct as a potentially active component of an intubation technique. It does not prove that stylet use caused the difference between MACMAN and DEVICE, or identify the best adjunct for every video blade.⁸

The 2022 Cochrane review broadened the perspective to 222 studies and 26,149 participants, mostly undergoing elective surgery. For Macintosh-style video devices, failed intubation was reduced (risk ratio 0.41; 95% CI 0.26–0.65), while the evidence for increased first-attempt success was rated low certainty (risk ratio 1.05; 95% CI 1.02–1.09). Other video designs also appeared to reduce failed intubation. Its search preceded DEVICE and only 21 studies took place outside operating theatres. Consequently, this synthesis supported the general promise of video laryngoscopy while leaving a setting-specific emergency and ICU question. DEVICE addressed that gap more directly; it did not make all pooled devices, populations and outcome definitions interchangeable.⁹

Taken together, the evidence available by the end of 2023 favours video laryngoscopy as the initial device in settings resembling DEVICE, particularly where trainees work with familiar equipment and appropriate tube-delivery aids. The remaining uncertainty is chiefly about the magnitude of benefit in other systems, the best device and adjunct combination, and patient-centred outcomes. The contrast between technical and physiological success is consistent across this literature. INTUBE's frequent cardiovascular events were observed over a longer window and with a different definition, so its complication rate cannot be compared directly with DEVICE's composite. Both nevertheless reinforce the need to assess airway mechanics and physiological preparation as related but distinct parts of care.^{1,5}

Should video laryngoscopy be the first choice for emergency intubation in critical illness?

Video laryngoscopy improves first-attempt success in settings comparable to DEVICE, with familiar equipment and routine tube-delivery aids. A reduction in severe complications or mortality remains unproven.

References

1. Russotto V, Myatra SN, Laffey JG, Tassistro E, Antolini L, Bauer P, et al.; INTUBE Study Investigators. Intubation Practices and Adverse Peri-intubation Events in Critically Ill Patients From 29 Countries. JAMA. 2021;325(12):1164–1172.

2. Prekker ME, Driver BE, Trent SA, Resnick-Ault D, Seitz K, Russell DW, et al.; DEVICE investigators and the Pragmatic Critical Care Research Group. DirEct versus Video LaryngosCope (DEVICE): protocol and statistical analysis plan for a randomised clinical trial in critically ill adults undergoing emergency tracheal intubation. *BMJ Open*. 2023;13(1):e068978.
3. Janz DR, Semler MW, Lentz RJ, Matthews DT, Assad TR, Norman BC, et al. Randomized Trial of Video Laryngoscopy for Endotracheal Intubation of Critically Ill Adults. *Crit Care Med*. 2016;44(11):1980–1987.
4. Lascarrou JB, Boissrame-Helms J, Bailly A, Le Thuaut A, Kamel T, Mercier E, et al.; Clinical Research in Intensive Care and Sepsis (CRICS) Group. Video Laryngoscopy vs Direct Laryngoscopy on Successful First-Pass Orotracheal Intubation Among ICU Patients: A Randomized Clinical Trial. *JAMA*. 2017;317(5):483–493.
5. Prekker ME, Driver BE, Trent SA, Resnick-Ault D, Seitz KP, Russell DW, et al.; DEVICE Investigators and the Pragmatic Critical Care Research Group. Video versus Direct Laryngoscopy for Tracheal Intubation of Critically Ill Adults. *N Engl J Med*. 2023;389(5):418–429.
6. DEVICE Investigators. Supplementary appendix to: Video versus Direct Laryngoscopy for Tracheal Intubation of Critically Ill Adults. *N Engl J Med*. 2023. doi:10.1056/NEJMoa2301601.
7. Driver BE, Prekker ME, Moore JC, Schick AL, Reardon RF, Miner JR. Direct Versus Video Laryngoscopy Using the C-MAC for Tracheal Intubation in the Emergency Department, a Randomized Controlled Trial. *Acad Emerg Med*. 2016;23(4):433–439.
8. Jaber S, Rollé A, Godet T, Terzi N, Riu B, Asfar P, et al.; STYLETO trial group. Effect of the use of an endotracheal tube and stylet versus an endotracheal tube alone on first-attempt intubation success: a multicentre, randomised clinical trial in 999 patients. *Intensive Care Med*. 2021;47(6):653–664.
9. Hansel J, Rogers AM, Lewis SR, Cook TM, Smith AF. Videolaryngoscopy versus direct laryngoscopy for adults undergoing tracheal intubation. *Cochrane Database Syst Rev*. 2022;4:CD011136.

AMBROISIE

Landais M, Nay MA, Auchabie J, Hubert N, Frerou A, Yehia A, et al.; REVA network and CRICS-TriggerSEP F-CRIN research network. Continued enteral nutrition until extubation compared with fasting before extubation in patients in the intensive care unit: an open-label, cluster-randomised, parallel-group, non-inferiority trial. Lancet Respir Med. 2023;11(4):319–328.

Introduction

Stopping gastric feeding before planned extubation is an intuitively appealing precaution. Removing the tracheal tube changes airway protection, and subsequent respiratory deterioration may require urgent reintubation. Reducing the stomach's contents might reduce regurgitation and aspiration around these events. However, the practical question is whether a prescribed interruption in feeding actually improves patient outcomes sufficiently to justify postponing extubation and withholding nutrition.¹

Practice has varied markedly. A 2009 survey of US training programmes found median fasting periods before extubation ranging from 1.5 hours in medical critical care programmes to six hours in anaesthesia programmes. Only 80 of 297 programmes responded, limiting representativeness, but the variation illustrates the absence of a shared evidential basis. A customary interval is not necessarily a validated physiological threshold, particularly when gastric emptying is unpredictable during critical illness.^{1,2}

The decision also has consequences beyond the volume of formula delivered. A patient judged ready for extubation may remain invasively ventilated while a fasting interval elapses. Conversely, removing that interval may advance extubation into an earlier phase of recovery. Any comparison must therefore examine the combined consequences of feeding policy, extubation timing and subsequent respiratory management, rather than assume that calories or gastric volume alone explain its results.¹

The Ambroisie Project addressed this uncertainty by allocating intensive care units (ICUs) to continue established enteral nutrition until extubation or to use fasting with gastric suction. Its non-inferiority hypothesis was that continued feeding would not cause an unacceptable increase in extubation failure. The trial therefore tested whether a precaution could be omitted within everyday ICU care; it did not seek to establish that uninterrupted feeding was superior for survival or that aspiration could never occur.^{1,3}

Synopsis

AMBROISIE was a pragmatic, open-label, parallel-group, cluster-randomised non-inferiority trial in 22 French ICUs. Recruitment ran from 1 April 2018 to 31 October 2019. Eleven units were allocated to each strategy before patient inclusion, using a computer-generated sequence prepared by an independent statistician and stratified by university status and predominant case mix. Unit allocation aimed to avoid contamination between feeding strategies. The French Ministry of Health funded the trial and had no role in its design, conduct, analysis, interpretation or publication decisions.¹

Eligible adults had received invasive ventilation for at least 48 hours and prepyloric enteral nutrition for at least 24 hours when the decision to extubate was made. Important exclusions included a decision against reintubation, tracheostomy, postpyloric feeding, previous enrolment, pregnancy and specified legal or social-protection restrictions. Patients or surrogates provided consent, with subsequent patient consent when initially impossible. Weaning, spontaneous breathing trials (SBTs) and assessment of readiness for extubation remained discretionary.^{1,4}

In intervention units, enteral nutrition continued at its existing rate until extubation, without gastric suction before, during or after the procedure. In comparator units, feeding stopped for at least six hours before extubation, with continuous gastric suction when the existing tube's diameter permitted it. The comparator thus combined fasting and attempted gastric emptying. The same allocated strategy applied to subsequent extubations after reintubation. Feeding after extubation, retention of the gastric tube, respiratory support and other care remained at clinicians' discretion.^{1,4}

The primary endpoint was extubation failure, defined as reintubation or death within seven days after extubation. Clinicians determined the need for reintubation without masking. Secondary outcomes included failure within 48 and 72 hours, pneumonia within 14 days while in the ICU, respiratory complications, glucose disturbances, nutritional delivery, workload and ICU mortality. Hypoglycaemia was defined as glucose below 0.7 g/L. An independent masked committee adjudicated pneumonia using radiological, clinical and microbiological information. An investigator's attribution of reintubation to aspiration was a separate outcome.^{1,4}

Assuming a 16% control failure rate, the non-inferiority margin was an absolute ten percentage points. The planned 1,100 patients across 22 clusters provided 80% power at a one-sided 0.025 significance level, allowing for an intracluster correlation of 0.03 and unequal cluster sizes. Non-inferiority required the upper limit of the two-sided 95%

confidence interval (CI) for continued feeding minus fasting to remain below ten percentage points. No interim analysis or missing-data imputation was planned.^{1,4}

Both intention-to-treat and per-protocol analyses were specified. Absolute differences incorporated clustering; mixed log-binomial models estimated risk ratios. Some secondary event analyses, including pneumonia, used Poisson models with ICU observation time as an offset and therefore produced rate ratios. Time to ICU discharge accounted for the competing risk of death. Post-hoc analyses explored adjustment for imbalanced patient characteristics and care.⁴

Of 4,198 patients assessed for eligibility, 1,130 were enrolled: 617 in continued-feeding units and 513 in fasting units. Another 1,168 eligible patients were not enrolled, most commonly because of investigator unavailability or other recorded reasons. The per-protocol population comprised 595 and 413 patients, respectively. Continued-feeding exclusions followed interruptions exceeding 30 minutes or reductions exceeding half the feeding rate; comparator exclusions predominantly involved interruption of gastric suction. Unintentional extubation before six hours of fasting did not itself require per-protocol exclusion.^{1,4}

Mean ages were 61.7 and 64.1 years, and most patients had medical admissions. Median ventilation duration was approximately eight days in both groups. Important differences accompanied unit allocation: corticosteroids specifically to prevent laryngeal oedema were given to 2.8% versus 31.4%, while weak or ineffective cough was more frequent in fasting units. Physiotherapy after extubation was more frequent in continued-feeding units. Prophylactic noninvasive ventilation (NIV) was used in 24.1% versus 27.1%, and high-flow nasal oxygen in 13.1% versus 17.9%.¹

The median enteral volume delivered during the six hours before extubation was 375 mL versus zero. Median gastric aspirate volume was zero versus 20 mL. In fasting units, the mean interval from stopping nutrition to extubation was 10.3 hours, and from stopping to resuming nutrition 22.7 hours. Among patients with timing data, the median interval from a successful SBT to extubation was 2.0 versus 17.6 hours. Recorded enteral and parenteral intake on the day of extubation was a median 880 versus 700 kcal; this was not a comparison of two prescribed daily caloric targets.¹

Seven-day extubation failure occurred in 106/617 patients (17.2%) with continued feeding and 90/513 (17.5%) with fasting. The absolute difference was -0.4 percentage points (95% CI -5.2 to 4.5), meeting the non-inferiority criterion; the risk ratio was 0.98 (95% CI 0.74–1.29). Eleven and nine patients, respectively, died without reintubation.

Per-protocol failure occurred in 101/595 (17.0%) versus 74/413 (17.9%), an absolute difference of -0.9 percentage points (95% CI -5.6 to 3.7), also meeting non-inferiority.¹

Earlier failure was numerically more common with continued feeding. At 48 hours, 77/617 (12.5%) versus 44/513 (8.6%) had failed extubation, an absolute difference of 3.9 percentage points (95% CI -0.1 to 7.9); the modelled risk ratio was 1.46 (95% CI 1.02–2.07). At 72 hours, the corresponding rates were 14.7% versus 10.5%, with an absolute difference of 4.2 percentage points (95% CI -0.1 to 8.6). A post-hoc covariate-adjusted analysis of the seven-day endpoint gave a difference of 1.7 percentage points (95% CI -1.5 to 4.9), remaining within the margin.^{1,4}

Pneumonia occurred in 10/617 patients (1.6%) versus 13/513 (2.5%); the rate ratio was 0.77 (95% CI 0.22–2.69). Reintubation attributed to aspiration occurred in eight versus 15 patients. Vomiting within one hour of extubation occurred in eight versus one. Before extubation on the procedure day, hypoglycaemia occurred in two versus 16 patients. These outcomes had different ascertainment methods and should not be combined into a single aspiration or safety endpoint.^{1,4}

ICU mortality was 24/617 (3.9%) versus 35/513 (6.8%). The absolute difference was -2.9 percentage points (95% CI -6.4 to 0.5), whereas the modelled risk ratio was 0.56 (95% CI 0.32–0.99). The median interval from successful SBT to ICU discharge was 4.0 versus 6.6 days. These secondary findings accompanied the primary non-inferiority result but did not establish a survival advantage.¹

Critique

AMBROISIE established non-inferiority against its chosen margin, not equality of risk. A ten-percentage-point margin would have allowed failure to rise from the assumed 16% to almost 26% while still potentially satisfying the design. That tolerance is clinically generous for an endpoint containing reintubation and death. The observed upper confidence limit of 4.5 percentage points is more reassuring than the margin itself, but still leaves a potentially important excess compatible with the data. Whether that residual uncertainty is acceptable depends on the value assigned to earlier extubation and avoiding interruption of feeding.¹

The comparator also matters to the interpretation of non-inferiority. Fasting with suction was a customary precaution whose protective benefit had not been established by the evidence presented for this trial. There was therefore no proven comparator effect to preserve. The useful result is an estimate of the consequences of two care policies, with a prespecified threshold for acceptable harm. Calling continued feeding

simply “as safe” would conceal both the chosen tolerance and the limited precision of uncommon complications.^{1,4}

Randomising units was sensible for an intervention embedded in routine nursing and extubation practice, and it reduced the likelihood of staff alternating inconsistently between strategies. However, 1,130 patients did not provide 1,130 independent treatment allocations: only 22 units were randomised. Statistical accommodation of clustering corrects the uncertainty associated with correlated observations; it cannot ensure that 11 units per group have matching extubation cultures. The substantial differences in steroid prophylaxis, cough assessment and physiotherapy show why this distinction is consequential.¹

Some differences could influence failure independently of feeding. More oedema prophylaxis in fasting units might favour the comparator, whereas more postextubation physiotherapy in continued-feeding units might favour the intervention; the observed trial cannot reliably assign or quantify either effect. Weaker cough in fasting units may reflect patient mix, assessment practice or both. Post-hoc adjustment supported the primary non-inferiority conclusion, but it depended on measured variables and modelling choices. Adjustment for care delivered after allocation also cannot recreate an individually randomised comparison with identical accompanying treatment.^{1,4}

Recruitment occurred after clinicians knew their unit's assignment, and more eligible patients were not enrolled than entered either individual group. This does not demonstrate selective recruitment, particularly when many omissions were attributed to availability, but it leaves an opportunity for selection linked to local practice. The absence of standardised readiness criteria further means that the study population was defined partly by each unit's decision to extubate. These limitations are more relevant than the simple observation that the trial was open label.^{1,4}

The large difference in SBT-to-extubation time is central, rather than an incidental process result. Avoiding delay is a plausible advantage of omitting fasting. It also means that the two groups were extubated at different points after readiness was first established. Higher failure at 48 and 72 hours, followed by similar seven-day cumulative failure, may reflect this timing difference or other aspects of care. It cannot establish that feeding caused early respiratory deterioration, nor should the early signal disappear from the appraisal because the primary endpoint was later.¹

Even the early estimates require care. Their absolute confidence intervals included zero, while the corresponding modelled risk-ratio intervals excluded one. Such differences can arise from the distinct methods used to estimate uncertainty; choosing only the more

striking result would exaggerate certainty. Numerous secondary comparisons further increase the possibility of chance findings. The defensible conclusion is a concerning early numerical pattern requiring interpretation alongside the seven-day result, without converting it into definitive proof of early harm.^{1,4}

The endpoint usefully counted death without reintubation as failure, avoiding classification of those patients as successful extubations. However, reintubation remained a clinical decision, susceptible to differing thresholds and use of rescue support. The publicly accessible protocol abstract described reintubation within seven days, whereas the final report explicitly included death. Without the full dated original protocol and amendment history, it is not possible to determine whether this difference reflects abbreviated abstract wording or a formal endpoint change. It warrants documentary verification, not an allegation of outcome switching.^{1,3}

Agreement between intention-to-treat and per-protocol analyses strengthens the margin-based finding. Nevertheless, the latter excluded 22 intervention patients and 100 comparator patients, largely because maintaining suction proved difficult. Removing patients according to delivered care can disrupt randomisation and select different populations in the two groups. Conversely, retaining unintentional early extubations was clinically coherent if such events were consequences of delaying a planned procedure. These choices illustrate why the intention-to-treat estimate remains the clearest description of adopting each policy.^{1,4}

The evidence for aspiration safety is less precise than the primary result. Pneumonia adjudication was masked, a concrete protection against differential diagnostic interpretation, but only 23 patients met that endpoint. Its rate-ratio interval ranged from a substantial reduction to a sizeable increase. Surveillance ended with ICU discharge, which occurred earlier in the continued-feeding group; accounting for observation time does not establish the absence of later pneumonia. Investigator-attributed aspiration leading to reintubation also misses aspiration not followed by that decision. Neither outcome proves that gastric contents were harmless.^{1,4}

The secondary mortality estimates likewise should not be promoted into a survival claim. Only 59 ICU deaths occurred, and the absolute and relative confidence intervals differed in whether they excluded no effect. Earlier ICU discharge is useful, but is also influenced by local discharge practice and the starting point used for measurement. The practical inference applies most directly to predominantly medical patients already receiving gastric feeding after several days of ventilation, with reintubation available. It does not establish a policy for postpyloric feeding, immediate oral intake, inability to protect the airway, or every surgical population.¹

Body of Evidence

The relevant earlier literature addressed several different questions: what clinicians customarily did, whether fasting emptied the stomach, whether feeding precautions prevented pneumonia, and how other extubation interventions affected failure. These questions are related but not interchangeable. Before AMBROISIE, a demonstration that practice varied or that a gastric measurement was imperfect could support a trial, but could not establish the clinical safety of continued feeding at tube removal.¹

Schneider and colleagues' 2009 survey described striking specialty differences in withholding enteral nutrition. Its low response rate limits an estimate of how common any particular interval was, and it collected institutional practice rather than comparative patient outcomes. It therefore provides historical evidence that fasting policy depended on the clinical setting, not evidence that the shortest or longest interval was preferable. AMBROISIE replaced this descriptive uncertainty with a randomised comparison of actual management policies.^{1,2}

The 2013 randomised trial by Reignier and colleagues examined another precaution: routine measurement of gastric residual volume during mechanical ventilation and early enteral feeding. Among 449 analysed patients in nine French ICUs, ventilator-associated pneumonia occurred in 38/227 (16.7%) without monitoring and 35/222 (15.8%) with monitoring. The absolute difference was 0.9 percentage points, with a 90% CI of -4.8 to 6.7, meeting that study's ten-percentage-point non-inferiority margin. Vomiting was more frequent without monitoring. This supported questioning an established feeding process while preserving attention to harms, but concerned feeding during ventilation rather than the transition through extubation.⁵

NUTRIGUS, published in 2021 after AMBROISIE had completed recruitment, examined gastric ultrasound immediately before extubation. Of 100 analysed patients, 26 had a stomach classified as full; fasting for at least six hours did not distinguish patients with an empty stomach from those fasting for less time. Among the 57 patients receiving enteral nutrition during the preceding 48 hours, fasting duration did not correlate with gastric antral area. These observational findings challenge the assumption that elapsed time reliably guarantees gastric emptying. They do not determine aspiration risk, validate ultrasound-guided extubation, or supply the clinical outcome comparison that AMBROISIE provides.⁶

Other randomised evidence explains why accompanying extubation care deserves scrutiny. In the 2007 trial by François and colleagues, methylprednisolone begun 12 hours before planned extubation reduced postextubation laryngeal oedema. Among 698 analysed patients, reintubation occurred in 13/355 receiving methylprednisolone and

26/343 receiving placebo. That study tested a timed drug regimen, not feeding. Nevertheless, it makes the large difference in oedema prophylaxis between AMBROISIE's groups a clinically plausible competing influence rather than a purely administrative imbalance. It cannot show how much that difference affected AMBROISIE's result.^{1,7}

HIGH-WEAN, published near the end of AMBROISIE recruitment in 2019, evaluated respiratory support after extubation in high-risk patients. High-flow nasal oxygen alternating with prophylactic NIV reduced seven-day reintubation compared with high-flow oxygen alone: 40/339 (11.8%) versus 55/302 (18.2%), an absolute difference of -6.4 percentage points (95% CI -12.0 to -0.9). This was a different, individually randomised intervention in a selected population. Its relevance is that extubation failure can be modified by management after tube removal, independently of the preceding feeding interval. It was not evidence available when AMBROISIE was designed.⁸

By the end of 2023, AMBROISIE therefore offered direct evidence to reconsider routine fasting with suction in patients resembling those enrolled. Its contribution was to connect a feeding policy with patient outcomes and an appreciable change in timing, rather than infer benefit from gastric volume alone. Continued feeding met the planned non-inferiority criterion, while clinically important uncertainty remained about smaller increases in failure and rare aspiration complications. Applying the result requires preserving the trial's distinction between continuing an established feed until extubation and independently deciding readiness, airway protection, respiratory support and nutrition afterwards.^{1,4}

Must established gastric feeding be stopped before planned extubation?

Routine fasting with suction can be reconsidered in patients resembling those enrolled. Continued feeding met non-inferiority, but did not exclude a clinically important smaller increase in failure or uncommon aspiration harm.

References

1. Landais M, Nay MA, Auchabie J, Hubert N, Frerou A, Yehia A, et al.; REVA network and CRICS-TriggerSEP F-CRIN research network. Continued enteral nutrition until extubation compared with fasting before extubation in patients in the intensive care unit: an open-label, cluster-randomised, parallel-group, non-inferiority trial. *Lancet Respir Med.* 2023;11(4):319–328.

2. Schneider JA, Lee YJ, Grubb WR, Denny J, Hunter C. Institutional practices of withholding enteral feeding from intubated patients. *Crit Care Med*. 2009;37(7):2299–2302.
3. Landais M, Nay MA, Auchabie J, Hubert N, Rebion A, Robert A, et al. Continuous enteral nutrition compared with a maximal gastric vacuity strategy at the time of extubation in the intensive care unit: protocol for a non-inferiority cluster randomised trial (the Ambroisie Project). *BMJ Open*. 2021;11(5):e041799.
4. Landais M, Nay MA, Auchabie J, Hubert N, Frerou A, Yehia A, et al. Supplementary appendix 2 to: Continued enteral nutrition until extubation compared with fasting before extubation in patients in the intensive care unit. *Lancet Respir Med*. 2023;11:319–328.
5. Reignier J, Mercier E, Le Gouge A, Boulain T, Desachy A, Bellec F, et al. Effect of Not Monitoring Residual Gastric Volume on Risk of Ventilator-Associated Pneumonia in Adults Receiving Mechanical Ventilation and Early Enteral Feeding: A Randomized Controlled Trial. *JAMA*. 2013;309(3):249–256.
6. Nguyen M, Drihem A, Berthoud V, Dransart-Raye O, Bartamian L, Gounot I, et al. Fasting does not guarantee empty stomach in the intensive care unit: A prospective ultrasonographic evaluation (The NUTRIGUS study). *Anaesth Crit Care Pain Med*. 2021;40(6):100975.
7. François B, Bellissant E, Gissot V, Desachy A, Normand S, Boulain T, et al. 12-h pretreatment with methylprednisolone versus placebo for prevention of postextubation laryngeal oedema: a randomised double-blind trial. *Lancet*. 2007;369(9567):1083–1089.
8. Thille AW, Muller G, Gacouin A, Coudroy R, Decavèle M, Sonnevile R, et al.; HIGH-WEAN Study Group and the REVA Research Network. Effect of Postextubation High-Flow Nasal Oxygen With Noninvasive Ventilation vs High-Flow Nasal Oxygen Alone on Reintubation Among Patients at High Risk of Extubation Failure: A Randomized Clinical Trial. *JAMA*. 2019;322(15):1465–1475.

REMICRUSH

Grillot N, Lebuffe G, Huet O, Lasocki S, Pichon X, Oudot M, et al.; Atlanrea Study Group and Société Française d'Anesthésie Réanimation (SFAR) Research Network. Effect of Remifentanyl vs Neuromuscular Blockers During Rapid Sequence Intubation on Successful Intubation Without Major Complications Among Patients at Risk of Aspiration: A Randomized Clinical Trial. JAMA. 2023;329(1):28–38.

Introduction

Rapid sequence induction aims to provide suitable conditions for tracheal intubation while limiting the interval during which an anaesthetised patient remains vulnerable to aspiration. The conventional combination of a hypnotic and a rapidly acting neuromuscular blocking drug achieves unconsciousness and muscle relaxation through different mechanisms. Removing the latter component is attractive only if the replacement preserves reliable intubating conditions and offers a worthwhile advantage. Avoiding paralysis cannot be assumed to make induction safer simply because one category of adverse drug effects has been removed.¹

The concerns motivating an alternative include anaphylaxis, prolonged neuromuscular blockade and, with particular drugs or susceptible patients, metabolic complications. Remifentanyl offers a different approach. This potent, rapidly acting opioid can suppress responses to airway instrumentation when combined with a hypnotic and has a short duration of action. However, it also depresses respiration and can produce hypotension and bradycardia. A short duration does not guarantee that spontaneous ventilation will return before a patient with failed intubation develops hypoxaemia. The relevant comparison therefore concerns the complete induction strategy, including the ease of obtaining the airway and the physiological consequences of doing so.¹

Small anaesthesia studies had shown that intubation without neuromuscular blockade was feasible, with performance dependent on opioid dose and accompanying anaesthesia. These observations did not establish that substituting remifentanyl was acceptably effective in patients selected specifically because they were at risk of aspiration. Nor could favourable assessments of intubating conditions establish a reduction in serious complications.^{2,3}

REMICRUSH tested whether remifentanyl plus a hypnotic was non-inferior to a rapid-onset neuromuscular blocker plus a hypnotic for successful intubation at the first attempt without major complications. Its setting was the operating room, and its

question was replacement of neuromuscular blockade for the first attempt, rather than the value of adding an opioid to an otherwise conventional induction.⁴

Synopsis

This investigator-initiated, open-label, parallel-group, randomised non-inferiority trial recruited patients in 15 French private and university hospitals between October 2019 and April 2021. The French Ministry of Health funded the study without involvement in its conduct or publication decisions. Nantes University Hospital, the sponsor, coordinated operations, monitoring and quality checks and conducted the statistical analyses. An independent data and safety monitoring board reviewed safety reports.⁴

Adults aged 18–80 years requiring orotracheal intubation for general anaesthesia were eligible with at least one specified aspiration risk: fasting for less than six hours, bowel obstruction, recent vomiting or orthopaedic trauma, or relevant gastro-oesophageal disease or dysfunction. The vomiting and trauma windows were 12 hours. Important exclusions included predicted difficult intubation, preoperative oxygen saturation below 95%, haemodynamic shock, cardiac arrest, pregnancy and contraindications preventing use of the study drugs. The final protocol defined shock as mean arterial pressure of 65 mmHg or less or vasopressor use. Written consent was obtained before participation according to the principal report.^{1,4}

Allocation occurred immediately before the procedure through a secure web system, using a 1:1 ratio and variable blocks of four or six. An independent methodologist generated the sequence. Stratification concerned bowel obstruction and the planned first laryngoscopy device: direct or video. Clinicians selected that device before allocation. Treating clinicians were unblinded. A research assistant or independent observer, separate from the intubating operator, collected procedural data prospectively; the report did not describe blinded outcome assessment.⁴

In the intervention arm, intravenous remifentanyl 3–4 µg/kg followed the hypnotic immediately, with laryngoscopy starting 30–60 seconds later and no neuromuscular blocker before the first attempt. Controls received succinylcholine 1 mg/kg or rocuronium 1 mg/kg after the hypnotic, without an opioid for induction before the first attempt. Succinylcholine was preferred; rocuronium was recommended when it was contraindicated. The final protocol used lean weight for remifentanyl and rocuronium in patients with body mass index of at least 30 kg/m², and actual weight for succinylcholine. The laryngoscopy interval was similar, with cessation of fasciculations guiding timing after succinylcholine.^{1,4}

Hypnotic choice and dose remained discretionary. Both groups underwent preoxygenation, described in the principal report as three minutes. Routine mask ventilation during apnoea, a stylet and cricoid pressure were not mandatory. Rescue mask ventilation and additional drugs were permitted for desaturation or unexpected difficulty; neuromuscular blockade could therefore be given after an unsuccessful remifentanyl attempt. Subsequent anaesthesia was not required to remain free of neuromuscular blockers.^{1,4}

The primary outcome required successful tube placement during the first laryngoscope insertion and freedom from six major complications during the first ten minutes after induction. These were operator-observed aspiration of liquid or solid material at the larynx, oxygen saturation below 95%, major haemodynamic instability, sustained arrhythmia, cardiac arrest and grade III–IV anaphylaxis. The haemodynamic definition used mean arterial pressure at or below 50 mmHg or at or above 110 mmHg for more than three minutes. Sustained arrhythmia required a new rhythm or rate disturbance lasting over 30 seconds and requiring treatment. The final protocol also incorporated rescue mask ventilation for desaturation within its desaturation criterion. Secondary outcomes included procedural difficulty and timing, recovery-room complications and pneumonia, acute respiratory distress syndrome and in-hospital mortality through day seven. Separate analyses of primary components were designated post hoc.^{1,4,5}

Assuming 80% success with neuromuscular blockade, 1,150 participants provided 80% power with a seven-percentage-point non-inferiority margin and one-sided α of 0.025. For the difference in success proportions, remifentanyl minus control, the lower two-sided 95% confidence limit had to remain above –7 percentage points. Both as-randomised and per-protocol analyses were required. Mixed-effects logistic regression adjusted for the two stratification factors and included centre as a random effect; adjusted absolute differences used Wald confidence intervals. Fourteen missing primary outcomes were multiply imputed across ten datasets using demographic and clinical variables. Secondary comparisons were not adjusted for multiplicity. No efficacy interim analysis was planned.^{1,4}

Of 1,154 recorded participants, four had cancelled surgery and 1,150 were randomised, 575 per arm. A complete screening denominator was not collected. Twenty participants were excluded from the per-protocol analysis for eligibility or consent restrictions, nonreceipt of allocated treatment or missing outcome data, leaving 565 per arm. Mean ages were 49.6 and 51.7 years; approximately half were women, four-fifths were American Society of Anesthesiologists class I–II, and 60% underwent urgent or semiurgent procedures. Baseline mean oxygen saturation was 97.8% in both groups.

Gastrointestinal obstruction, ileus or vomiting was common, but this was predominantly a physiologically stable surgical population.⁴

Allocated treatment was received by 574 and 573 participants. Propofol was used in approximately 98%, with mean total doses of 246 and 236 mg; a February 2023 correction replaced erroneous mg/kg labels in the original table. Mean remifentanyl dose was 241 µg. Among controls with agent data, 408/570 received succinylcholine and 156/570 rocuronium. Direct laryngoscopy was used in approximately 85%; stylets or bougies in 11.1% and 15.8%. Bag-mask preoxygenation predominated, while ventilation during apnoea was recorded in approximately 1%.^{4,6}

Among participants with observed primary outcomes, success occurred in 374/566 (66.1%) versus 408/570 (71.6%). The adjusted as-randomised analysis, including imputation, gave a difference of –6.1 percentage points (95% CI –11.6 to –0.5; P=0.37 for non-inferiority). Per-protocol success was 374/565 (66.2%) versus 403/565 (71.3%), with an adjusted difference of –5.7 points (95% CI –11.3 to –0.1; P=0.32). Neither analysis established non-inferiority. Both intervals excluded zero in favour of neuromuscular blockade.⁴

First-attempt success considered alone was 502/567 (88.5%) versus 533/570 (93.5%); the adjusted difference was –4.9 points (95% CI –8.2 to –1.6). Desaturation below 95% occurred in 42/567 versus 38/570, and observed aspiration in four versus two patients. Major haemodynamic instability, prolonged arrhythmia or cardiac arrest occurred in 118/567 (20.8%) versus 111/570 (19.5%); the adjusted difference was 1.1 points (95% CI –3.7 to 5.9). No cardiac arrests occurred. Mean induction-to-intubation time, measured to the sixth capnography waveform, was 2.5 minutes in both groups. Rescue neuromuscular blockade was used in 19/568 versus 8/570.⁴

Vasopressors after induction were recorded in 360/568 (63.4%) versus 180/570 (31.6%), with a reported adjusted difference of 35.7 points (95% CI 30.2 to 41.1). Severe adverse events were reported as 12/575 (2.1%) versus 3/575 (0.5%), adjusted difference 1.8 points (95% CI 0.4 to 3.2). By day seven, pneumonia occurred in 3/569 versus 2/571 and in-hospital death in one participant per arm. Prespecified subgroup analyses identified no significant interaction; exploratory analyses suggested variation between centres, without a detectable temporal trend.^{4,5}

Critique

REMICRUSH provides more information than a simple failure to prove non-inferiority. The confidence interval for its primary comparison lay wholly below zero, and the

per-protocol analysis agreed. Under the reported models, remifentanyl produced fewer successful first-attempt intubations without major complications. At the same time, the intervals crossed the -7-point margin. They therefore include losses smaller than the prespecified unacceptable difference and losses larger than it. Statistical inferiority was demonstrated, but inferiority exceeding the clinical margin was not. Describing the trial as merely inconclusive would obscure the directional evidence; describing it as proving a clinically unacceptable decrement would overstate its precision.⁴

The margin itself requires scrutiny. Allowing success to fall from an anticipated 80% to 73% tolerates seven additional composite failures per 100 patients. That trade-off could be reasonable only alongside a compensating benefit valued by patients or clinicians. REMICRUSH did not demonstrate such a benefit. Avoiding a neuromuscular blocker at induction did not produce a clear reduction in aspiration, important postoperative complications or serious adverse events. The point estimate fell within the permitted loss, but a point estimate alone cannot establish non-inferiority. The confidence interval is the relevant safeguard against adopting a materially less effective strategy.^{1,4}

The composite combines procedural success with complications of markedly different severity. A second laryngoscope insertion, transient desaturation below 95%, sustained severe blood-pressure disturbance and cardiac arrest do not have equal consequences. The difference was principally associated with first-attempt performance, rather than a demonstrated excess of aspiration or death. That remains clinically relevant in patients at risk of aspiration, but the study does not quantify how often the additional attempts caused subsequent serious injury. Conversely, low counts of aspiration and arrest cannot establish comparable safety. The primary result should not be translated into a six-percentage-point increase in catastrophic complications.⁴

Procedural observations support a plausible explanation for the difference. Vocal cords were recorded as abducted in 489/557 patients receiving remifentanyl and 548/560 controls, and coughing requiring increased anaesthesia occurred more frequently with remifentanyl. These findings fit less favourable intubating conditions without paralysis. They are also susceptible to observer judgement, and their exploratory status precludes treating them as separately confirmed mechanisms. The objective requirement for successful tube placement is more persuasive than a subjective difficulty score, although operators determine when to stop an attempt or give rescue drugs.^{4,5}

Allocation concealment, selection of the initial device before randomisation and prospective collection by an observer separate from the operator protect the comparison. Operators were nevertheless unblinded, with succinylcholine-associated fasciculations making concealment difficult. Clinicians could alter anaesthetic dosing,

haemodynamic support and rescue decisions in response to allocation. Stylet or bougie use also differed. These are concrete routes for performance effects, not simply an abstract objection to an open trial. The study estimates the effect of implementing the two strategies with those accompanying decisions; it cannot isolate the pharmacological effect of remifentanyl while holding every other intervention constant.⁴

The large difference in vasopressor use is especially relevant. It prevents a reassuring interpretation based solely on similar frequencies of the primary haemodynamic component. Treatment can prevent a blood-pressure threshold from being reached or shorten its duration. However, the available results do not resolve the timing and indications sufficiently to establish whether every administration treated hypotension or anticipated it. The independently recorded vascular adverse events also favoured controls. Together these observations weaken the proposed haemodynamic advantage, without proving that all additional support represented harmful instability caused by remifentanyl.^{1,5}

The dose and timing define the intervention tested. REMICRUSH used a bolus of 3–4 µg/kg after the hypnotic, followed by an early attempt, predominantly with propofol. It did not test a titrated infusion, an opioid-first sequence or opioid supplementation of neuromuscular blockade. A post hoc comparison of lower and higher remifentanyl doses gave an adjusted success difference of –4.6 points (95% CI –12.7 to 3.4), but dose was selected clinically rather than randomised. It cannot establish that a higher fixed dose would remove the disadvantage. Any improvement in intubating conditions would also need assessment against respiratory and cardiovascular effects.^{1,4}

The documentary chronology deserves attention. The original protocol allowed up to two laryngoscopies within the primary definition; a November 2019 amendment changed this to fewer than two and described it as correction of a typing error to match the original registry entry. Recruitment had already begun. This explanation should be acknowledged without asserting that an outcome was changed after examination of treatment results. The final SAP was dated November 2020, during recruitment, and specified centre adjustment, both principal analysis populations and multiple imputation. The supplied English translations were prepared in 2022. These documents permit scrutiny, but do not establish the investigators' access to accumulating comparative outcomes at each amendment.¹

Missing primary data were uncommon, and the consistent as-randomised and per-protocol findings reduce concern that the conclusion depends on one population definition. They do not make imputed observations equivalent to measured ones. The report's Figure 2 supplies the observed denominators of 566 and 570, whereas the

abstract and Table 3 attach the same percentages to 575 per arm. Separating observed counts from the imputed analysis avoids reproducing this inconsistency.⁴

The protocol's withdrawal section specified excluding remifentanyl patients given rescue neuromuscular blockade from the per-protocol analysis. The final SAP's population definition did not repeat that rule, and the published per-protocol table retained 19 rescue recipients. Agreement between analyses is supportive, but the per-protocol result does not represent adherence to every exclusion instruction in the supplied protocol.^{1,5}

Applicability is narrower than the general phrase "patients at risk of aspiration" suggests. Hypoxaemia, shock and predicted difficult intubation were excluded; most participants had little systemic disease. The trial therefore cannot validate remifentanyl as an alternative in the unstable patient or the anticipated difficult airway. It also excluded patients unable to receive the comparator drugs, a group in whom an alternative might be most attractive. Low use of video laryngoscopy and airway adjuncts further defines the setting. Subgroup confidence intervals were too wide to establish equivalence with video laryngoscopy, while exploratory centre variation cannot identify a reproducibly successful local technique.^{1,4}

Finally, follow-up was short and clinically important postoperative events were rare. The protocol assumed patients discharged before day seven were free of complications, limiting ascertainment after discharge. Neuromuscular blockers and reversal could be used subsequently in the remifentanyl arm, so later respiratory outcomes do not compare complete avoidance with exposure throughout anaesthesia. The defensible clinical inference concerns routine replacement for initial intubation: effectiveness was reduced under the studied conditions, and a compensating safety advantage was not established.^{1,4,5}

Body of Evidence

Before REMICRUSH, small studies established feasibility and explored dose, while leaving a substantial gap between acceptable intubating conditions and reliable protection of an at-risk patient. Their procedures, drug sequences and outcomes differed sufficiently that apparently favourable results could not simply be transferred to rapid sequence induction. The original abstracts of the following early studies support their principal findings; they do not permit a complete independent assessment of allocation, exclusions or outcome adjudication.

In a 1999 comparison by Alexander and colleagues, 60 premedicated patients of American Society of Anesthesiologists class I–II received propofol 2 mg/kg followed by remifentanyl 2 µg/kg, alfentanil 50 µg/kg or succinylcholine 1 mg/kg. At intubation one minute later, good or excellent conditions occurred in 35%, 85% and 100%, respectively. This was an early warning that remifentanyl at that dose was an unreliable substitute. Its relevance lies in dose selection and procedural conditions, rather than an estimate of serious complications in patients with aspiration risk.⁷

A separate 1999 dosing study from the same group enrolled 60 class I–II patients receiving propofol 2 mg/kg and remifentanyl 3, 4 or 5 µg/kg. Good or excellent conditions were recorded in 12 patients at 3 µg/kg and 19 patients in each higher-dose group. Blood pressure and heart rate fell after induction in all three groups. These observations supplied a rationale for investigating higher doses, but their small scale did not establish that improved conditions outweighed physiological disadvantages in a broader surgical population.²

Erhan and colleagues subsequently randomised 80 healthy patients to alfentanil 40 µg/kg or remifentanyl 2, 3 or 4 µg/kg before propofol 2 mg/kg. Excellent conditions were reported in 45%, 20%, 75% and 95%, respectively. The opioid was administered over 90 seconds and intubation attempted 90 seconds after propofol, a different sequence and timescale from REMICRUSH. The study supported dose-dependent suppression of airway responses; it did not test non-inferiority to a neuromuscular blocker in patients selected for aspiration risk. Its findings explain the later hypothesis without resolving it.³

The 2017 Cochrane review placed such studies within a wider evidence base of 34 randomised trials and 3,565 participants. Across all trials, avoiding neuromuscular blockade increased the reported risk of difficult intubation (risk ratio 5.00, 95% CI 3.49–7.15), with substantial heterogeneity. Most participants were relatively healthy and many trials excluded difficult airways. The review also cautioned that study definitions of difficulty did not necessarily represent a life-threatening airway problem. Thus, the direction of evidence favoured blockade, while its clinical translation and certainty remained limited. REMICRUSH addressed an important population gap with a substantially larger individual trial.⁸

CURASMUR, published in December 2019 after REMICRUSH recruitment began, addressed the choice between rapid-onset neuromuscular blockers in out-of-hospital emergency intubation. In its per-protocol analysis, first-attempt success was 455/610 with rocuronium 1.2 mg/kg and 489/616 with succinylcholine 1 mg/kg: a difference of –4.8 points, with a one-sided 97.5% confidence interval extending from –9.0 to infinity. Rocuronium did not meet the seven-point non-inferiority margin. This different setting,

higher rocuronium dose and procedural endpoint prevent a direct cross-trial ranking. It also means that REMICRUSH's pooled comparator cannot determine the effect of remifentanyl against each blocker separately.⁹

By the end of 2023, the evidence therefore supported continued neuromuscular blockade for routine rapid sequence induction in patients resembling those enrolled in REMICRUSH. Earlier demonstrations that intubation was possible without paralysis had not established an equally dependable strategy. REMICRUSH strengthened the case for preserving first-attempt effectiveness, while leaving rare safety outcomes and exceptional circumstances unresolved. It provided no basis for inferring that avoiding paralysis would improve safety in the hypoxaemic or shocked patient, and it did not answer whether an opioid is useful as an adjunct to neuromuscular blockade.⁴

Should remifentanyl replace neuromuscular blockade for rapid sequence intubation in patients at risk of aspiration?

Routine replacement is not supported in patients resembling those studied. Remifentanyl failed non-inferiority, reduced first-attempt success and offered no demonstrated safety advantage.

References

1. Grillo N, Lebuffe G, Huet O, Lasocki S, Pichon X, Oudot M, et al. Trial protocol, amendments and statistical analysis plans for REMICRUSH. Supplement 1 to: Effect of Remifentanyl vs Neuromuscular Blockers During Rapid Sequence Intubation. JAMA. 2023;329(1):28–38.
2. Alexander R, Olufolabi AJ, Booth J, El-Moalem HE, Glass PS. Dosing study of remifentanyl and propofol for tracheal intubation without the use of muscle relaxants. Anaesthesia. 1999;54(11):1037–1040.
3. Erhan E, Ugur G, Alper I, Gunusen I, Ozyar B. Tracheal intubation without muscle relaxants: remifentanyl or alfentanil in combination with propofol. Eur J Anaesthesiol. 2003;20(1):37–43.
4. Grillo N, Lebuffe G, Huet O, Lasocki S, Pichon X, Oudot M, et al.; Atlanrea Study Group and Société Française d'Anesthésie Réanimation (SFAR) Research Network. Effect of Remifentanyl vs Neuromuscular Blockers During Rapid Sequence Intubation on Successful Intubation Without Major Complications Among Patients at Risk of Aspiration: A Randomized Clinical Trial. JAMA. 2023;329(1):28–38.

5. Grillo N, Lebuffe G, Huet O, Lasocki S, Pichon X, Oudot M, et al. Supplementary online content for REMICRUSH: additional figures, tables and data safety report. Supplement 2 to: Effect of Remifentanyl vs Neuromuscular Blockers During Rapid Sequence Intubation. JAMA. 2023;329(1):28–38.
6. Incorrect Units of Measure in a Table. JAMA. 2023;329(6):515.
7. Alexander R, Booth J, Olufolabi AJ, El-Moalem HE, Glass PS. Comparison of remifentanyl with alfentanil or suxamethonium following propofol anaesthesia for tracheal intubation. Anaesthesia. 1999;54(11):1032–1036.
8. Lundstrøm LH, Duez CHV, Nørskov AK, Rosenstock CV, Thomsen JL, Møller A, et al. Avoidance versus use of neuromuscular blocking agents for improving conditions during tracheal intubation or direct laryngoscopy in adults and adolescents. Cochrane Database Syst Rev. 2017;5(5):CD009237.
9. Guihard B, Chollet-Xémard C, Lakhnati P, Vivien B, Broche C, Savary D, et al. Effect of Rocuronium vs Succinylcholine on Endotracheal Intubation Success Rate Among Patients Undergoing Out-of-Hospital Rapid Sequence Intubation: A Randomized Clinical Trial. JAMA. 2019;322(23):2303–2312.



Respiratory Trials

AMIKINHAL

Ehrmann S, Barbier F, Demiselle J, Quenot JP, Herbrecht JE, Roux D, et al.; Reva and CRICS-TRIGGERSEP F-CRIN Research Networks. Inhaled Amikacin to Prevent Ventilator-Associated Pneumonia. N Engl J Med. 2023;389(22):2052–2062.

Introduction

Preventing ventilator-associated pneumonia (VAP) requires more than preventing bacterial entry into the lung. Clinicians must also minimise exposure to invasive ventilation, support airway clearance and avoid preventive treatments whose harms outweigh the infections they avert. By 2022, guidance emphasised minimising sedation, assessing readiness for extubation, mobilisation, head elevation and oral care. Yet patients requiring prolonged ventilation remained vulnerable despite these measures. The question was whether an additional, short antimicrobial intervention could interrupt the development of infection without imposing a substantial burden of toxicity or resistance.¹

Inhaled amikacin offers a plausible way to target this process. Delivery through the ventilator circuit can expose the tracheobronchial tree and endotracheal-tube biofilm to antibiotic concentrations difficult to achieve with systemic treatment. Administering the drug before established pneumonia might also avoid some of the problems of delivering aerosol to poorly aerated, infected lung. However, deposition depends on the device, circuit and ventilation, and nebulisation can cause bronchospasm or obstruct expiratory filters. Limited systemic absorption does not establish either clinical safety or ecological neutrality.^{2,3}

Earlier trials of preventive inhaled antibiotics produced inconsistent results. A double-blind trauma trial of aerosolised ceftazidime did not reduce VAP, while a small trial of nebulised colistin suggested benefit for some infection outcomes without establishing a difference in its primary endpoint. These studies left uncertainty about the antibiotic, population and treatment window most likely to benefit.^{4,5}

AMIKINHAL tested a deliberately limited strategy: three daily doses of nebulised amikacin beginning after three days of invasive ventilation. This timing aimed to identify patients with continuing exposure to VAP risk while avoiding treatment of those extubated early. The central hypothesis was that this brief course would reduce subsequent pneumonia; the broader clinical question was whether preventing that diagnosis would translate into better recovery.²

Synopsis

AMIKINHAL was an investigator-initiated, parallel-group, double-blind superiority trial conducted in 19 French intensive care units. Recruitment ran from 3 July 2017 to 9 March 2021. The regional university hospital in Tours conducted the trial, with funding from the French Ministry of Health and additional association support. An independent data and safety monitoring committee reviewed outcomes. Adults were eligible after at least 72 hours of invasive ventilation through an endotracheal tube, provided randomisation occurred within 96 hours.²

Patients were excluded for suspected or confirmed VAP, expected extubation within 24 hours, concurrent systemic aminoglycoside treatment, tracheostomy, or relevant drug contraindications. Renal exclusions included stage 2 or 3 acute kidney injury without renal replacement therapy and chronic kidney disease with estimated glomerular filtration below 30 mL/min. Patients receiving, or scheduled to receive, renal replacement therapy could participate despite acute kidney injury. Written consent was obtained from patients or proxies. During pandemic visiting restrictions, an approved arrangement allowed oral proxy consent followed by written confirmation.^{2,3}

Central web-based randomisation assigned patients equally to amikacin or placebo, stratified by centre and systemic antibiotic use on the day of allocation. An independent statistician generated the concealed sequence using blocks of four. Patients and clinicians were blinded. Staff uninvolved in patient care prepared the solutions, and opaque equipment concealed their appearance. Routine serum amikacin measurements were prohibited to protect blinding. The primary outcome was adjudicated by an independent blinded committee.²

The intervention was amikacin 20 mg/kg of ideal body weight once daily for three consecutive days; controls received an equivalent volume of 0.9% sodium chloride. Both solutions were delivered through an Aerogen Solo vibrating-mesh nebuliser in the inspiratory limb. The appendix describes preparation at 125 mg/mL, opaque tubing and syringe delivery, expiratory filters, and a bedside safety checklist. Ventilator settings, sedation, neuromuscular blockade and management of heated humidification remained at clinician discretion. This was an adjunct to other care, including systemic antibiotics and VAP prevention measures. Subsequent doses were withheld after extubation, qualifying acute kidney injury or a clinical requirement for systemic aminoglycosides.^{2,3}

The primary endpoint was the first VAP episode between randomisation and day 28. Diagnosis required compatible clinical and radiographic findings together with a positive quantitative pulmonary culture. The diagnostic pathway used a new infiltrate with at least two clinical features involving abnormal white-cell count, fever or purulent

secretions; deterioration in patients with acute respiratory distress syndrome could prompt investigation despite difficulty identifying a new infiltrate. Investigators submitted suspected episodes, clinical and microbiological information, and chest radiographs for adjudication.^{2,3}

Secondary outcomes included VAP caused by amikacin-susceptible Gram-negative bacteria, VAP incidence density, ventilator-associated events, systemic antibiotic exposure, ventilation duration, length of stay and mortality. The ventilator-associated event framework separately identified sustained deterioration in oxygen or positive end-expiratory pressure requirements, infection-related complications and possible pneumonia. Antibiotic exposure was measured both as days receiving at least one systemic antibiotic and as the sum of individual antibiotic-days. Safety assessment included respiratory events, kidney injury and resistant organisms.^{2,3}

A sample of 850 patients provided 80% power to detect a reduction in VAP from 12% to 6%, using a two-sided 5% significance threshold and a competing-risk framework. The publication states that the proportional-hazards assumption failed on evaluation of the data. The primary comparison therefore used the difference in restricted mean time to VAP, accounting for death and extubation as competing events. Other time-to-event analyses used Fine–Gray models; count outcomes used quasi-Poisson regression with exposure-time offsets. Secondary comparisons were not adjusted for multiplicity and were exploratory. Missing secondary data were not imputed.^{2,3}

Of 6,419 patients screened after more than 48 hours of ventilation, 850 were randomised: 420 to amikacin and 430 to placebo. Three amikacin patients withdrew consent, leaving 417 and 430 in the reported intention-to-treat population. There were no missing primary-outcome data in this population. Mean age was approximately 61–62 years, about two-thirds were men, and most admissions were medical. Admission Simplified Acute Physiology Score II averaged 52 in each group. Allocation occurred a median of 84 hours after intubation, and 78% versus 77% were receiving systemic antibiotics.^{2,3}

At least one nebulisation was delivered to 412 amikacin and 428 placebo patients; all three were delivered to 337 and 355, respectively. Extubation was the commonest reason for an incomplete course. Mean nebulisation duration was 47 versus 49 minutes. At the first treatment, approximately half used a dry circuit and half a circuit with heated humidification; most humidifiers were left active. Volume-assist control and pressure-support ventilation accounted for almost all treatments. These details demonstrate delivery across more than one ventilatory configuration.^{2,3}

VAP occurred in 62/417 patients receiving amikacin (15%) and 95/430 receiving placebo (22%). The reported difference in restricted mean time to VAP was 1.5 days (95% confidence interval [CI] 0.6–2.5; $P=0.004$). First-episode incidence density was 16 versus 23 per 1,000 invasive-ventilation days, rate ratio 0.68 (95% CI 0.49–0.94). VAP due to susceptible Gram-negative bacteria occurred in 31 versus 61 patients, with a restricted-mean difference of 1.9 days (95% CI 1.1–2.8).²

Ventilator-associated conditions occurred in 137 patients (33%) versus 170 (40%), hazard ratio 0.79 (95% CI 0.64–0.99). Infection-related complications occurred in 74 (18%) versus 111 (26%), hazard ratio 0.66 (95% CI 0.50–0.89). Neither systemic antibiotic exposure measure demonstrated a clear reduction. Among patients discharged alive, median invasive-ventilation duration was nine days in both groups; ICU stay was 12 versus 13 days and hospital stay 27 days in both. Day-28 deaths numbered 92 versus 106, with a reported difference of –3 percentage points (95% CI –9 to 3); 12 patients lacked day-28 mortality data. At day 90, deaths numbered 134 versus 143, with 16 patients missing and a reported difference of –2 percentage points (95% CI –8 to 5).^{2,3}

Serious adverse events considered related to the trial occurred in seven amikacin patients (1.7%) and four placebo patients (0.9%), principally involving airway or circuit complications. Among patients without kidney injury at allocation, subsequent injury occurred in 11/294 versus 24/309, hazard ratio 0.47 (95% CI 0.23–0.96). Acquired resistance to amikacin was identified in 41 patients in each group. Analyses restricted to patients receiving all three doses favoured amikacin, while the smaller baseline-colonisation and tracheobronchitis subgroups had imprecise estimates.^{2,3}

Critique

AMIKINHAL establishes a reduction in adjudicated VAP under the tested conditions more convincingly than it establishes an improvement in patients' overall course. The difference between 15% and 22% is clinically substantial for an infection endpoint. Nevertheless, the trial did not demonstrate shorter ventilation, less systemic antibiotic exposure or better survival. These findings should be presented together. It would be unreasonable to dismiss a prevented infection as meaningless, but equally unreasonable to assume that the associated downstream benefits have already been established.^{2,3}

The primary endpoint is particularly sensitive to the mechanism of the intervention. Inhaled antibiotics may prevent infection, reduce bacterial burden sufficiently to change culture positivity, or do both. Because microbiological confirmation formed part of the VAP definition, a treatment effect on the diagnostic process cannot be separated completely from an effect on clinically important pneumonia. Investigators submitted

120 suspected episodes with amikacin and 144 with placebo; adjudication confirmed 52% and 66%, respectively. The reduction in suspected episodes was less convincing than the adjudicated comparison. This does not invalidate adjudication, but makes corroborating clinical outcomes especially valuable.³

Several design features strengthen confidence. Allocation was concealed, treatment appearance was masked, and a separate blinded committee reviewed the primary endpoint. Nebulised saline gave clinicians a similar procedure in both groups, reducing the opportunity for differential diagnostic attention or co-interventions. No primary-outcome data were missing among the 847 analysed patients. These protections matter because chest-radiograph interpretation, secretion assessment and decisions to investigate pneumonia contain subjective elements. The three post-randomisation exclusions mean the reported population was not literally every randomised patient, but their small number is unlikely by itself to explain the observed separation.^{2,3}

The alternative ventilator-associated event definitions offer partial corroboration. Infection-related complications were less frequent with amikacin, and the broader ventilator-associated condition endpoint also favoured treatment. Their additional physiological requirements make the findings useful, but they are overlapping secondary outcomes from the same patients rather than independent replications. Infection-related definitions also incorporate antibiotic decisions and microbiology. Without multiplicity adjustment, their confidence intervals describe uncertainty for individual comparisons; they should not be treated as a collection of separately proven benefits.^{2,3}

The statistical presentation needs careful translation. Death and extubation alter the opportunity to develop VAP and must be addressed as competing events. The reported 1.5-day restricted-mean difference belongs to that VAP analysis. It is not 1.5 additional days alive, a 1.5-day reduction in ventilation, or a delay that every treated patient experienced. The paper explains that failure of proportional hazards motivated the change in analytical approach. That is a methodological rationale, although the publications alone do not provide a complete independent audit of analytical prespecification. A post hoc Fine–Gray analysis beginning on day 7 also favoured amikacin, but changing the time window cannot replace the principal analysis.^{2,3}

The crude absolute difference, calculated from 62/417 and 95/430, is approximately 7.2 percentage points, corresponding to about 14 patients treated per additional patient avoiding the observed primary endpoint. This is a descriptive calculation, not the trial's modelled treatment effect or a competing-risk-adjusted number needed to treat. The placebo event rate was substantially above the anticipated 12%, whereas the relative

reduction was smaller than the 50% sought. These differences caution against assuming the same absolute benefit in a unit with a much lower underlying VAP incidence.²

Treatment delivery supports practical feasibility but defines the intervention narrowly. A three-day course after 72 hours avoids exposing patients who need only brief ventilation. Approximately four-fifths completed all doses, with many early discontinuations explained by extubation rather than failure of implementation. Conversely, a per-protocol comparison conditioned on completing treatment selects patients according to events after randomisation, including continued ventilation and renal function. It is supportive, not a superior estimate of the effect of assignment. Similarly, analyses by circuit configuration were not randomised comparisons of humidification strategies and cannot establish which configuration is best.^{2,3}

The absence of clear differences in recovery outcomes has several possible explanations, including limited power, competing causes of prolonged critical illness and a smaller clinical effect than the diagnostic endpoint suggests. Survivor-only ventilation and length-of-stay comparisons also condition on survival, which occurs after randomisation. Post hoc analyses including nonsurvivors answer a different question because an early death can shorten treatment duration. The appendix's observational comparison of patients with and without VAP cannot resolve this: prolonged ventilation itself increases infection risk, and patients developing VAP differ from those who do not. Association between pneumonia and a longer stay does not quantify the stay that amikacin prevents.³

Safety remains an active part of the decision. Airway and expiratory-filter complications occurred despite trained staff and a limited course. Saline nebulisation exposes controls to procedural risk, so the comparison does not estimate the additional procedural harm of introducing nebulisation where none would otherwise occur. The favourable kidney-injury result is an exploratory finding in a selected renal population, not evidence that amikacin protects the kidney. Likewise, identical counts of acquired amikacin resistance over short individual follow-up do not exclude longer-term effects on patients, unit ecology or transmission.^{2,3}

Applicability is strongest for adults still ventilated through an endotracheal tube at three to four days, without pneumonia and without the specified renal or aminoglycoside exclusions. It is weaker for tracheostomised patients, those extubated early, settings with different resistance patterns, and centres without reliable aerosol delivery and filter monitoring. Most patients were already receiving systemic antibiotics, so this was not a test of inhaled prophylaxis in an otherwise antibiotic-unexposed population. The evidence supports considering the tested strategy in an appropriate prevention

programme, while leaving universal implementation and net patient benefit unresolved.^{2,3}

Body of Evidence

The preceding literature did not establish a uniform effect of preventive inhaled antibiotics. In the 2007 trial by Claridge and colleagues, 105 high-risk trauma patients were randomised to seven days of aerosolised ceftazidime or placebo. VAP at two weeks occurred in 21/53 ceftazidime patients and 24/52 controls; by 30 days the figures were 26/53 and 26/52. The published abstract did not identify a difference in multidrug-resistant infections or other infectious complications. This directly relevant prevention trial provided little support for routine ceftazidime in that population. Its much smaller size and trauma-specific setting limit comparison with AMIKINHAL, but its findings discourage treating all inhaled antibiotics as interchangeable.⁴

Karvouniaris and colleagues subsequently randomised 168 patients in a Greek ICU with prevalent multidrug-resistant organisms to nebulised colistin or saline. Treatment was given three times daily for up to ten ICU days or extubation. The trial was open-label, unlike AMIKINHAL. Its primary 30-day VAP incidence was 16.7% versus 29.8% ($P=0.07$). VAP incidence density and some microbiological secondary outcomes favoured colistin, but those findings did not turn the primary comparison into a definitive demonstration of benefit. The numerical direction was compatible with AMIKINHAL, while the drug, treatment frequency, duration and ecological setting differed. Together, the trials suggest a plausible prevention effect that depends on the specific strategy rather than a settled class effect.⁵

Treatment of established pneumonia asks a different question. INHALE, published before the AMIKINHAL results, evaluated inhaled amikacin added to systemic antibiotics in 153 ICUs across 25 countries. It randomised 725 patients to 400 mg every 12 hours for ten days or placebo. The primary analysis included 508 patients meeting its microbiological and severity criteria and receiving treatment. Survival at days 28–32 was 191/255 with amikacin and 196/253 with placebo, odds ratio 0.841 (95% CI 0.554–1.277; $P=0.43$). This substantial negative treatment trial does not disprove short preventive therapy before VAP develops. It does, however, demonstrate why high local antibiotic delivery and microbiological plausibility cannot substitute for testing patient outcomes.⁶

AMIKINHAL therefore contributes a distinct comparison: a larger blinded prevention trial, using a brief course of amikacin in patients selected by continued ventilation rather than confirmed bacterial infection. Its result should neither be diluted by indiscriminately combining unlike aerosol strategies nor insulated from the wider lesson that

infection-related endpoints and recovery can diverge. Drug selection, timing, dose, nebuliser and background antibiotic exposure are parts of the intervention. The available trials do not establish whether another aminoglycoside, a longer course or culture-targeted prophylaxis would reproduce its effect.^{2,4,5,6}

The practical reference point remains a broader prevention programme. The 2022 guidance prioritised reducing avoidable ventilation and sedation, daily assessment for liberation, mobilisation, head elevation and oral care. These measures address risk through pathways that differ from airway antibiotic delivery. AMIKINHAL tested an additional intervention against a background of prevention practice; it did not compare antibiotic prophylaxis with more consistent application of those measures. Local implementation would therefore require understanding both baseline VAP burden and the capacity to deliver and monitor nebulisation.^{1,2}

By the end of 2023, AMIKINHAL had materially strengthened evidence that a brief preventive inhaled-antibiotic strategy can reduce adjudicated VAP. The next evidential step was to determine whether the effect produces sufficient patient benefit to justify wider exposure, workload and ecological uncertainty. Trials focused on clinically important recovery, antibiotic consumption and sustained resistance surveillance would answer that question more directly than additional demonstrations of airway bacterial suppression. Until then, the supported conclusion is specific: fewer VAP diagnoses with this regimen in this selected ventilated population, with the overall balance for routine adoption still requiring judgement.^{2,3}

Should inhaled amikacin be used to prevent VAP after three days of ventilation?

A three-day course reduces adjudicated VAP in selected patients. Benefits for survival, ventilation duration and systemic antibiotic use remain unproven, so routine use requires careful consideration.

References

1. Klompas M, Branson R, Cawcutt K, Crist M, Eichenwald EC, Greene LR, et al. Strategies to prevent ventilator-associated pneumonia, ventilator-associated events, and nonventilator hospital-acquired pneumonia in acute-care hospitals: 2022 Update. *Infect Control Hosp Epidemiol.* 2022;43(6):687–713.

2. Ehrmann S, Barbier F, Demiselle J, Quenot JP, Herbrecht JE, Roux D, et al.; Reva and CRICS-TRIGGERSEP F-CRIN Research Networks. Inhaled Amikacin to Prevent Ventilator-Associated Pneumonia. *N Engl J Med*. 2023;389(22):2052–2062.
3. Ehrmann S, Barbier F, Demiselle J, et al. Supplementary Appendix to: Inhaled Amikacin to Prevent Ventilator-Associated Pneumonia. *N Engl J Med*. 2023. doi:10.1056/NEJMoa2310307.
4. Claridge JA, Edwards NM, Swanson J, Fabian TC, Weinberg JA, Wood C, et al. Aerosolized ceftazidime prophylaxis against ventilator-associated pneumonia in high-risk trauma patients: results of a double-blind randomized study. *Surg Infect (Larchmt)*. 2007;8(1):83–90.
5. Karvouniaris M, Makris D, Zygoulis P, Triantaris A, Xitsas S, Mantzarlis K, et al. Nebulised colistin for ventilator-associated pneumonia prevention. *Eur Respir J*. 2015;46(6):1732–1739.
6. Niederman MS, Alder J, Bassetti M, Boateng F, Cao B, Corkery K, et al. Inhaled amikacin adjunctive to intravenous standard-of-care antibiotics in mechanically ventilated patients with Gram-negative pneumonia (INHALE): a double-blind, randomised, placebo-controlled, phase 3, superiority trial. *Lancet Infect Dis*. 2020;20(3):330–340.
7. AMIKINHAL investigators. Trial protocol and statistical analysis plan. Supplementary material to: Inhaled Amikacin to Prevent Ventilator-Associated Pneumonia. *N Engl J Med*. 2023. doi:10.1056/NEJMoa2310307.

Possible Reporting Inconsistencies. Figure 2 appears to mislabel follow-up as days since intubation. The Methods and supplied protocol define the primary pneumonia window from randomisation to day 28; the final SAP does not change that time origin. The randomisation clock is used in this chapter.^{2,7}

SiVent

Albert RK, Jurkovich GJ, Connett J, Helgeson ES, Keniston A, Voelker H, et al. Sigh Ventilation in Patients With Trauma: The SiVent Randomized Clinical Trial. JAMA. 2023;330(20):1982–1990.

Introduction

Mechanical ventilation after major trauma supports patients through several overlapping problems: impaired consciousness, shock, painful chest injury and direct pulmonary damage. Ventilation may be necessary even when gas exchange is initially well preserved. A strategy intended to prevent subsequent lung injury therefore addresses a different question from treatment of established acute respiratory distress syndrome (ARDS). SiVent examined whether periodically adding a larger breath to otherwise usual ventilation could improve recovery in this early period.¹

Spontaneous breathing includes occasional sighs. These may help maintain lung volume and the function of surfactant, the material that reduces surface tension within alveoli. Sedation and controlled ventilation interrupt that natural pattern. Reintroducing intermittent inflation offers a plausible physiological intervention, but the pressure and volume needed to produce a sigh may also overdistend lung or impair circulation. The relevant balance depends on the dose, frequency, underlying lung and haemodynamic reserve.¹

The broader evidence creates a tension. Low tidal volume ventilation improves important outcomes in established ARDS, making repeated exposure to larger volumes a legitimate concern. However, a brief intermittent breath is not equivalent to continuously ventilating with a high tidal volume. Small physiological studies demonstrated improved oxygenation during sigh delivery, and a later pilot randomised trial established feasibility during pressure support ventilation. Neither established that sighs improve survival or shorten ventilation in patients admitted with trauma.^{2,3,4}

SiVent tested a relatively infrequent, pressure-targeted sigh added early to usual care. Its rationale emphasised preserving surfactant function rather than using sustained recruitment to correct severe hypoxaemia. The clinical question was whether this simple intervention would increase days alive and free from ventilatory assistance, without introducing an unacceptable burden of adverse effects. That question required a patient-centred trial: improved oxygenation alone would not establish that patients recover sooner or survive more often.¹

Synopsis

SiVent was an unblinded, parallel-group randomised trial at 15 US academic trauma centres, recruiting from April 2016 to September 2022. The US Department of Defense funded the study and had no role in its design, conduct, analysis or publication. Adults admitted with trauma were eligible within 24 hours of starting mechanical ventilation if ventilation was expected to continue beyond 24 hours and survival beyond 48 hours. They also needed traumatic brain injury, more than one long-bone fracture, initial systolic pressure below 90 mmHg, lung contusion, or receipt of more than six units of blood products during the first 24 hours.¹

Important exclusions included inability to obtain consent, clinician unwillingness to enrol, pregnancy, imprisonment, burns exceeding 40% of body surface area, neurological impairment preventing spontaneous ventilation and irreversible disease associated with a greater than 50% estimated six-month mortality. Patients receiving airway pressure release ventilation were excluded. The required Dräger ventilator had to be available, and clinicians had to be willing to use low tidal volumes if ARDS developed. Written consent came from patients or their legal representatives. A central interactive web system allocated treatment equally, using randomly varying blocks of two, four or six, stratified by centre.¹

Both groups received clinician-directed ventilation, including the choice of mode, respiratory rate, tidal volume, positive end-expiratory pressure (PEEP) and inspiratory flow. The intervention added a five-second sigh every six minutes. Its volume was individually determined to achieve a plateau pressure of 35 cmH₂O, increased to 40 cmH₂O for body mass index above 35 kg/m² or presumed raised intra-abdominal pressure. Sighs continued during assisted ventilation but were not delivered during transport or surgery. The comparison was therefore an addition to heterogeneous usual care, rather than two fully standardised ventilation regimens.¹

The primary outcome was ventilator-free days to day 28 after randomisation, measured after the last successful liberation from assistance. Patients had to survive at least two calendar days after liberation and remain free from assistance. Death before day 28, or transition to comfort care, resulted in zero ventilator-free days. Reintubation generally reset the count, except for brief ventilation lasting less than 24 hours for surgery. Qualifying unassisted breathing could include oxygen or low-level continuous positive airway pressure without pressure support. Ten patients known to be alive at day 28, but with unknown ventilation status after transfer, were assigned zero days in a post hoc decision.^{1,5}

Secondary outcomes included 28-day mortality, ICU-free days, complications and discharge status. Complications were assessed by treating clinicians. Exploratory analyses included time to extubation, sedative exposure, oxygenation and radiographic surrogates of ARDS. A post hoc total ventilator-free-day measure counted qualifying periods without assistance before subsequent ventilation or death, unlike the primary measure based on final successful liberation.^{1,5}

The original sample size was 916, providing 90% power to detect a 2.25-day difference, assuming a standard deviation of 10.5 days and a two-sided significance level of .05. Following an interim review, the calculation used a standard deviation of 9.9 days, 80% power and a 1% withdrawal allowance, reducing the target to 544. Two interim reviews used a Lan-DeMets approach with a Pocock-family spending function. Recruitment ultimately stopped at 524 because funding ended, without a recommendation to stop for efficacy or futility.^{1,5}

The primary comparison used the Wilcoxon rank-sum test. Linear models also estimated unadjusted and adjusted mean differences. Adjustment included age, sex, smoking, the qualifying injury and shock characteristics, injury severity, baseline oxygenation category and centre. Missing injury-severity and oxygenation covariates were multiply imputed; a complete-case analysis was also reported. Mortality analyses used logistic and time-to-event models. Extubation analyses accounted for competing death. Secondary and subgroup comparisons were not adjusted for multiplicity.^{1,5}

Of 5753 patients screened, 524 were randomised: 261 to sighs and 263 to usual care. The primary analysis included 259 and 261 after three consent withdrawals and one patient with unknown vital status. All randomised patients contributed to time-to-event analyses with censoring as appropriate. Mean age was approximately 44 years, three-quarters were men and approximately 62% had traumatic brain injury. Mean injury-severity scores were 29–30. Baseline mean $\text{PaO}_2/\text{FiO}_2$ was approximately 349 mmHg in each group; more than half of those measured had values above 300 mmHg. Patients had received ventilation for approximately 17 hours before allocation.¹

Baseline tidal volumes averaged 7.4 versus 7.3 mL/kg predicted body weight, with PEEP approximately 7 cmH₂O. Sighs were delivered for a median of three days. Among 223 patients with recorded initial sigh settings, mean sigh volume was 939 mL, or 13.7 mL/kg predicted body weight. Four control patients temporarily received sighs. Recorded ordinary tidal volumes and sedation scores during follow-up were broadly similar between groups, although these measurements did not provide continuous assessment of all ventilation exposure.^{1,5}

Median ventilator-free days were 18.4 (interquartile range 7.0–25.2) with sighs and 16.1 (1.1–24.4) with usual care; the primary rank-sum comparison gave $P=.08$. The unadjusted mean difference was 1.9 days (95% confidence interval 0.1–3.6; $P=.04$), while the adjusted mean difference was 1.4 days (–0.2 to 3.0; $P=.08$). Among survivors, the adjusted mean difference was 0.8 days (–0.8 to 2.3). Complete-case and alternative ventilator-free-day analyses did not establish a consistent statistically persuasive benefit.^{1,5}

By day 28, 30 of 259 patients (11.6%) had died with sighs versus 46 of 261 (17.6%) with usual care: odds ratio 0.61 (95% confidence interval 0.37–1.00; $P=.05$). The adjusted mortality hazard ratio was 0.70 (0.43–1.15; $P=.16$). Brain injury accounted for 20 and 30 deaths respectively. Median ICU-free days were 13.7 versus 11.9; the adjusted mean difference was 0.8 days (–0.6 to 2.3). An exploratory extubation analysis gave a subdistribution hazard ratio of 1.21 (1.00–1.47; $P=.05$). No convincing treatment interaction emerged from the post hoc subgroups.^{1,5}

Non-fatal serious adverse events affected 80 patients in each group, approximately 31%, comprising 118 versus 111 events. Pneumothorax occurred in 18 versus 20 patients. Seven hypotension events occurred with sighs and none with usual care; five were temporally associated with sigh delivery. Four of these five patients were already receiving vasopressors. Ventilator-associated pneumonia occurred in 57 versus 69 patients. Sedative use was common in both groups, with a median seven days of exposure.¹

Critique

SiVent provides a clinically interesting signal, but it does not establish that routine sigh ventilation improves recovery after trauma. The principal difficulty is the combination of an inconclusive primary analysis, favourable alternative estimates and a borderline secondary mortality comparison. These findings should be considered together. Selecting the smallest P value would overstate certainty; treating every estimate as showing no possible benefit would also discard information. The adjusted primary interval includes both little effect and an improvement that could matter to patients and services.¹

The distinction between statistical comparisons is central. The rank-sum test compares outcome distributions, whereas the linear models estimate differences in mean ventilator-free days. Neither the unadjusted 1.9-day mean difference nor its confidence interval describes the difference between the reported medians. Skewness and the concentration of deaths at zero make these analyses sensitive to different features of

the data. A nominally significant unadjusted mean estimate cannot replace the designated primary rank comparison, particularly when the adjusted estimate crosses no effect. The appropriate interpretation is uncertainty about magnitude and robustness, rather than a contradiction requiring one result to be discarded.¹

Recruitment further limits precision. Funding termination left the trial 20 patients short of its revised target, but the larger change was the earlier reduction from 916 to 544 and from 90% to 80% power. That revision made the study less able to distinguish the planned difference from chance. The reported interim process did not identify an efficacy or futility reason for stopping, so this is not the usual problem of stopping at an unusually favourable interim result. Nevertheless, the final sample cannot answer the original question with its original precision. Failure to reach conventional significance does not establish equivalence or futility.^{1,5}

Ventilator-free days appropriately penalise death, but their interpretation requires examining both survival and liberation. The survivor-only difference was smaller than the overall difference, suggesting that mortality contributed to the favourable composite result. This does not prove that sighs primarily improve survival: restricting analysis to survivors selects patients according to a post-randomisation event and breaks the original comparability. Nor does the composite establish faster liberation in patients who would survive either treatment. The exploratory extubation estimate is compatible with that possibility, but its lower confidence limit is close to no effect and it belongs to a broader set of secondary analyses.¹

The mortality observation deserves confirmation rather than dismissal or immediate adoption. The absolute group proportions differ appreciably, yet the odds-ratio interval reaches 1.00 and the adjusted time-to-event interval is wide. Multiple outcomes were examined without multiplicity correction. Furthermore, brain injury accounted for most deaths, while few deaths were attributed to respiratory failure. A pulmonary intervention might have systemic consequences, but these data do not demonstrate the proposed pathway. No surfactant measurements or other direct mechanistic findings establish that the difference arose from preservation of surfactant function.¹

Missingness is small in number but conceptually important. Three withdrawals and one unknown vital status explain why the primary analysis contains fewer patients than randomised. Separately, ten patients alive at day 28 had unknown ventilation status and were assigned zero ventilator-free days. This is an assumption about their outcome, not observation that they remained ventilated throughout follow-up. Its effect depends on allocation and their unobserved recovery. Covariate imputation addresses a different issue and cannot resolve missing primary-outcome components. The report therefore

supports an analysis close to intention to treat, with stated exclusions and assumptions, rather than a completely observed outcome for every randomised participant.^{1,5}

Central randomisation and centre stratification provide a sound basis for comparison, but treatment was visible. Clinicians controlled ventilation, sedation and liberation decisions, all of which can influence ventilator-free days. Similar recorded sedation and ordinary tidal volumes are reassuring, although intermittent summaries cannot exclude differences in weaning readiness assessments or other care. Mortality is less vulnerable to assessment bias than a clinician-managed liberation endpoint, but decisions about treatment limitation may still matter in a population with frequent brain injury. There is no evidence here to claim that such decisions differed by allocation.^{1,5}

The intervention is more specific than the label “sigh ventilation” suggests. A five-second breath every six minutes, targeted to plateau pressure, differs from frequent sighs or prolonged recruitment manoeuvres. Initial delivered volumes were approximately twice ordinary tidal volumes, but pressure targeting produced variable volumes. Initial settings were available for 223 patients, which documents typical dose without establishing complete adherence across every scheduled breath. Median treatment lasted only three days. Benefits or harms from different frequencies, pressures, durations or ventilator implementations cannot be assumed to follow the same pattern.¹

Safety also warrants a qualified interpretation. Similar overall serious-event proportions and pneumothorax counts provide reassurance against a large common excess, but they do not establish physiological neutrality. Hypotension temporally related to sighs is consistent with a clinically plausible adverse effect of increased intrathoracic pressure. Several affected patients already required circulatory support; that vulnerability makes the observation relevant rather than dismissible. In a cohort with frequent traumatic brain injury, avoiding hypotension is especially consequential. The trial does not provide a basis for claiming a specific effect on intracranial pressure or cerebral recovery.¹

Generalisability is bounded by selection. Only a minority of screened patients entered the trial, with consent, clinician agreement and equipment availability affecting enrolment. Participants were predominantly younger men with trauma, often ventilated for neurological or other extrapulmonary reasons; most did not have severe oxygenation impairment at baseline. These features fit an early preventive hypothesis but limit extrapolation to established severe ARDS, older medical patients or prolonged ventilation. The findings support a further adequately powered trial of this defined intervention, with mortality, liberation and haemodynamic safety considered explicitly. They do not yet justify routine adoption across ventilated populations.¹

Body of Evidence

The development of sigh ventilation illustrates why a physiological response and a clinical treatment effect need separate evaluation. In 1999, Pelosi and colleagues studied ten patients with ARDS during sequential periods with and without sighs. Three consecutive sighs each minute, reaching a plateau pressure of 45 cmH₂O, improved oxygenation and end-expiratory lung volume during a one-hour intervention. Mean PaO₂ increased from approximately 93 to 138 mmHg, then returned towards baseline after sighs stopped. This supported short-term recruitment or stabilisation, but the small physiological study could not establish survival, liberation or uncommon harms. Its inflation frequency and pressure were also substantially greater than those tested by SiVent.³

The ARDS Network trial subsequently established the importance of ordinary tidal-volume exposure. Among 861 patients with acute lung injury or ARDS, ventilation targeting 6 rather than 12 mL/kg predicted body weight, alongside different plateau-pressure limits, reduced mortality from 39.8% to 31.0% and increased mean ventilator-free days from 10 to 12. The trial was stopped early for benefit. This evidence concerns sustained ventilation strategy, not occasional sighs. It supplies a reason to examine cumulative mechanical exposure carefully, while leaving open whether infrequent larger breaths can add benefit to lower-volume background ventilation.²

The ART trial demonstrated the danger of assuming that opening lung units necessarily improves outcomes. In 1010 patients with moderate or severe ARDS, a recruitment manoeuvre followed by compliance-guided PEEP titration increased 28-day mortality compared with low PEEP: 55.3% versus 49.3%, hazard ratio 1.20 (95% confidence interval 1.01–1.42). Pneumothorax and barotrauma also increased. ART tested a considerably more intensive recruitment-and-PEEP package, not SiVent's brief intermittent breaths. It therefore cannot supply a direct estimate of sigh-related harm. Its relevance is the broader requirement to demonstrate patient benefit and consider circulatory and mechanical costs, rather than use improved aeration as sufficient justification.⁶

PROTECTION provided a closer randomised comparison. This pilot enrolled 258 patients with acute hypoxaemic respiratory failure or ARDS receiving pressure support in 20 hospitals across eight countries. A three-second sigh at 30 cmH₂O once per minute was added to pressure support. The primary objective was non-inferiority for feasibility, assessed through failure of assisted ventilation, with a five-percentage-point margin. Failure occurred in 23% with sighs versus 30% without them, an absolute difference of –7 percentage points (95% confidence interval –18 to 4), meeting that objective. Mortality was 16% versus 21%, without a clear difference, and median ventilator-free days were 22 in both groups. These findings supported feasibility, but did not establish

clinical efficacy. Patients were already receiving assisted ventilation for hypoxaemic disease, and the sigh frequency was six times that used in SiVent.⁴

SiVent extends the question into early trauma care, including patients without marked hypoxaemia and patients receiving modes other than pressure support. Its design is consequently not a direct replication of PROTECTION. The intervention begins at a different stage, addresses a broader range of reasons for ventilation and uses a different dose. A favourable direction in both trials is encouraging but cannot be treated as independent confirmation of one identical treatment effect. In particular, a feasibility trial that meets a non-inferiority criterion does not become an efficacy trial because a later study has a suggestive mortality result.^{1,4}

By the end of 2023, this evidence supported a credible but unresolved hypothesis. Physiological studies showed that sighs can change lung function; pilot randomisation showed that a particular regimen can be delivered during pressure support; SiVent supplied a larger clinical comparison in trauma. None established a reproducible patient-centred benefit sufficient to define a routine regimen. The next question is not simply whether the ventilator can deliver sighs, but whether a specified pressure, frequency and timing improve survival or meaningful recovery in a clearly defined population without clinically important haemodynamic harm.¹

Should we add routine sigh breaths to early mechanical ventilation after major trauma?

Not routinely. The primary result was inconclusive, and the possible survival benefit requires confirmation.

References

1. Albert RK, Jurkovich GJ, Connett J, Helgeson ES, Keniston A, Voelker H, et al. Sigh Ventilation in Patients With Trauma: The SiVent Randomized Clinical Trial. *JAMA*. 2023;330(20):1982–1990.
2. Acute Respiratory Distress Syndrome 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(18):1301–1308.
3. Pelosi P, Cadringher P, Bottino N, Panigada M, Carrieri F, Riva E, et al. Sigh in acute respiratory distress syndrome. *Am J Respir Crit Care Med*. 1999;159(3):872–880.

4. Mauri T, Foti G, Fornari C, Grasselli G, Pinciroli R, Lovisari F, et al.; PROTECTION Trial Collaborators. Sigh in Patients With Acute Hypoxemic Respiratory Failure and ARDS: The PROTECTION Pilot Randomized Clinical Trial. *Chest*. 2021;159(4):1426–1436.
5. Albert RK, Jurkovich GJ, Connett J, Helgeson ES, Keniston A, Voelker H, et al. Supplementary online content to: Sigh Ventilation in Patients With Trauma: The SiVent Randomized Clinical Trial. *JAMA*. 2023.
6. Writing Group for the Alveolar Recruitment for Acute Respiratory Distress Syndrome Trial (ART) Investigators. Effect of Lung Recruitment and Titrated Positive End-Expiratory Pressure (PEEP) vs Low PEEP on Mortality in Patients With Acute Respiratory Distress Syndrome: A Randomized Clinical Trial. *JAMA*. 2017;318(14):1335–1345.

EXTUB-OBESE

De Jong A, Bignon A, Stephan F, Godet T, Constantin JM, Asehnoune K, et al.; EXTUB-OBESE trial group. Effect of non-invasive ventilation after extubation in critically ill patients with obesity in France: a multicentre, unblinded, pragmatic randomised clinical trial. *Lancet Respir Med.* 2023;11(6):530–539.

Introduction

Extubation removes both an artificial airway and the positive pressure that may have helped maintain lung volume. In patients with obesity, reduced functional residual capacity and increased respiratory loading can make that transition difficult. An apparently successful separation from invasive ventilation may be followed by atelectasis, impaired gas exchange or respiratory fatigue. Preventing these problems is attractive, especially when reintubation would expose the patient to another period of invasive support.¹

Non-invasive ventilation (NIV) supplies inspiratory assistance and positive end-expiratory pressure through a mask. It may support ventilation and limit alveolar collapse, but it also introduces discomfort, leaks, gastric distension and demands on staff. High-flow nasal oxygen (HFNO) offers a different balance of oxygen delivery, humidification and tolerance, with much less positive pressure. Standard oxygen remains a simpler option. The decision is therefore about the appropriate intensity of support, not merely which device produces the highest oxygen saturation.^{1,2}

Earlier evidence suggested that preventive NIV could benefit selected patients after extubation. An observational study in severe obesity found less respiratory failure than in historical controls. A large trial in patients at high risk of extubation failure supported NIV alternating with HFNO, and a subsequent exploratory analysis suggested a particularly favourable effect in patients with overweight or obesity. However, that subgroup was not an independent obesity-specific randomised trial, and the relevant populations and treatment durations varied.^{3,4,5}

EXTUB-OBESE addressed this uncertainty prospectively in a broad population with obesity, including medical and surgical patients. It compared immediate intermittent NIV, with oxygen between sessions, against oxygen alone, and separately randomised the oxygen-delivery method. The hypothesis was that early NIV would reduce failure of the assigned postextubation strategy. Whether that strategy-level benefit would also

mean fewer reintubations, shorter admission or improved survival was the more consequential clinical question.¹

Synopsis

EXTUB-OBESE was a pragmatic, multicentre, parallel-group randomised trial in 39 French ICUs. Recruitment ran from 2 October 2019 to 17 July 2021. Funding came from the French Ministry of Health, without funder involvement in design, analysis, interpretation or reporting. Adults with body mass index (BMI) at least 30 kg/m² on the day of extubation were eligible after more than six hours of invasive ventilation, when their treating clinician considered them ready for extubation. Written patient or proxy consent was sought, with provision for subsequent consent after emergency inclusion.^{1,2}

Patients were excluded for pre-extubation PaCO₂ at least 50 mmHg, isolated cardiogenic pulmonary oedema, tracheostomy, home NIV or continuous positive airway pressure, a decision not to reintubate, or anatomical barriers to NIV or HFNO. Pregnancy, breastfeeding and relevant legal or insurance restrictions also precluded enrolment. Arterial blood gases were not mandatory before extubation, and a spontaneous breathing trial was left to clinical discretion. The trial therefore did not require a uniform physiological assessment of readiness.^{1,2}

Two consecutive central allocations were performed. The first assigned NIV or oxygen therapy in a 1:1 ratio; the second assigned HFNO or standard oxygen within each group. Allocation used a computer-generated concealed sequence, with minimisation balancing centre, medical versus surgical admission and ventilation duration below versus at least 48 hours. Patients, clinicians and outcome assessors could not be blinded, but statisticians analysed masked treatment labels. The investigators described the oxygen allocation as nested within the main comparison, rather than a factorial design.¹

The first NIV session was offered within 30 minutes of extubation. Recommended PEEP was 10 cmH₂O, adjustable between 5 and 10 according to tolerance. Pressure support targeted respiratory rate 20–30 breaths/min and expired tidal volume 6–8 mL/kg predicted body weight. Sessions lasted 30–60 minutes, distributed across day and night, for at least four cumulative hours during the first 24 hours, without an upper limit. Weaning could begin after 24 hours if respiratory rate was stable below 25/min, PaO₂/FiO₂ exceeded 200 mmHg and PaCO₂ was below 45 mmHg.^{1,2}

Between NIV sessions, patients received their allocated oxygen method. HFNO was prescribed at 50 L/min for the first 24 hours, with inspired oxygen adjusted to maintain peripheral saturation at least 94%. Standard oxygen was given when saturation was 94%

or lower, using an interface appropriate to flow requirements. Oxygen support continued beyond 24 hours when required. Rescue NIV was permitted for patients assigned oxygen who developed respiratory failure, at their physician's request. Improving patients could step down from NIV without that being classified as treatment failure. Clinicians retained the final reintubation decision, guided by specified emergencies and signs of serious deterioration.^{1,2}

The primary endpoint was treatment failure within three days of extubation: reintubation, rescue switching to the other study treatment, or premature discontinuation. Each patient contributed only one component, prioritising reintubation, then switching without subsequent reintubation, then discontinuation. Premature discontinuation included failure to complete a 30-minute NIV session or 12 hours of oxygen because of patient request or medical reasons. The supplementary analysis description counted patients who died as reintubated failures. The main secondary endpoint was acute respiratory failure within seven days, requiring at least two specified clinical or physiological abnormalities. Later reintubation, oxygenation, organ dysfunction, lengths of stay and mortality were exploratory outcomes.^{1,2}

Assuming treatment failure in 12% with oxygen and 6% with NIV, 954 patients provided 80% power at a two-sided alpha of .05. The target increased to 1000 for loss to follow-up and consent withdrawal. The primary analysis compared all randomised groups using an uncorrected chi-squared test. Logistic and Cox models provided additional estimates, including centre and selected covariates. A prespecified per-protocol analysis excluded eligibility violations and reintubation for procedures without respiratory failure. Exploratory P values were adjusted using the Benjamini–Hochberg false-discovery-rate method. No primary outcome was missing; other missing data were not imputed. There was no data monitoring committee.^{1,2}

Of 1650 screened patients, 981 were enrolled: 490 assigned NIV and 491 oxygen. Each main group was divided approximately equally between HFNO and standard oxygen. The most frequent exclusion was home ventilatory support, affecting 284 screened patients. Mean age was 61 years and mean BMI approximately 35.5 kg/m². There were 585 surgical and 396 medical patients; approximately 12% had COVID-19. Median ventilation duration before extubation was 53 versus 56 hours overall, but approximately 17–21 hours in surgical patients and 144–164 hours in medical patients. A spontaneous breathing trial was recorded in 67% versus 70% of patients with available information.^{1,2}

Delivered NIV used almost exclusively an oronasal mask. Median PEEP was 6 cmH₂O and pressure support 7 cmH₂O. Median NIV exposure was five hours during the first 24 hours and seven hours during the first 72 hours. HFNO flow was a median 50 L/min; standard

oxygen flow was 4 L/min. Seventy oxygen-group patients received rescue NIV, a mean 29 hours after randomisation; 67 were not subsequently reintubated within the primary window.^{1,2}

Treatment failure occurred in 66 of 490 patients (13.5%) with NIV and 130 of 491 (26.5%) with oxygen: absolute difference –13.0 percentage points (95% confidence interval –17.9 to –8.1; $P<.0001$). The adjusted odds ratio was 0.42 (0.30–0.59). The composite comprised 48 versus 59 reintubations, zero versus 67 rescue switches without subsequent reintubation, and 18 versus four premature discontinuations. The three-day reintubation difference was –2.2 percentage points (–6.1 to 1.7; $P=.26$). No convincing interaction was found by oxygen device, admission type, ventilation duration or COVID-19 status.¹

Acute respiratory failure occurred in 54 of 489 versus 70 of 490 patients: difference –3.2 percentage points (95% confidence interval –7.4 to 0.92; $P=.13$). At day 28, reintubation had occurred in 87 of 489 versus 105 of 490, without a clear difference. Ninety-day mortality was 50 of 486 versus 52 of 490, and lengths of stay were similar. No death or cardiac arrest occurred during the interventions. Per-protocol reintubation was 31 of 466 versus 49 of 473 ($P=.042$). A post hoc analysis reassigned the 70 rescue-NIV recipients to the NIV group, producing reintubation rates of 51 of 560 versus 56 of 421 ($P=.037$).^{1,2}

Critique

The primary result establishes that starting NIV reduced failure of the assigned treatment strategy, but its clinical meaning is narrower than the large absolute difference initially suggests. Almost the entire numerical separation arose from rescue NIV in the oxygen group. Avoiding escalation is a legitimate practical objective, yet escalation is not equivalent to reintubation or death. The randomised evidence does not establish that giving NIV to everyone improves those more consequential outcomes.¹

The composite's asymmetry is fundamental. A patient assigned oxygen could deteriorate, receive NIV and recover without intubation, while still being classified as a primary failure. A patient assigned NIV could receive the same support and recover without meeting that component. Switching from NIV to oxygen for improvement was correctly not counted as failure, but the reverse rescue pathway had no clinically equivalent counterpart. This does not invalidate the chosen endpoint; it explains why the endpoint answers whether initial support sufficed, rather than whether the patient ultimately fared better.^{1,2}

Its components also differ in importance and susceptibility to judgement. Reintubation usually represents a major event. A change in device or a prematurely terminated session can range from consequential deterioration to intolerance of a relatively brief preventive treatment. The hierarchy avoids counting one patient several times, but it does not make these events interchangeable. NIV increased premature discontinuation, while reducing the switching component. Describing approximately eight patients treated to prevent one composite failure would therefore not mean eight treated to prevent one reintubation.¹

Treatment visibility increases this interpretative problem. Clinicians knew allocation and decided when rescue support was warranted; outcome assessors were also unblinded. Explicit recording of reasons for switching helps distinguish deterioration from routine de-escalation, but it cannot remove differences in clinical thresholds. Objective emergencies guided reintubation, while the broader management of deterioration remained discretionary. The primary result may contain both a physiological effect and the predictable consequences of beginning with a more intensive treatment. The trial cannot quantify their separate contributions.^{1,2}

The intention-to-treat reintubation comparison remains the most reliable estimate of the two starting strategies as implemented, including permitted rescue. Its confidence interval allows a clinically useful reduction but also little benefit or a small increase. The trial was powered for a six-point composite difference, not for modest differences in reintubation or mortality. Consequently, an inconclusive reintubation result should not be called proof that NIV is ineffective. Equally, the precision of the composite estimate cannot be transferred to survival or other patient-centred outcomes.¹

The post hoc crossover analysis does not repair this uncertainty. Moving patients according to treatment received abandons randomisation and assigns most successful rescue episodes to the NIV group. These patients were selected after their clinical course had begun; their prognosis, reasons for escalation and time exposed to oxygen differed from those initially assigned NIV. The analysis also combines immediate prevention with delayed rescue, which are different interventions. A lower reintubation proportion after this reassignment cannot establish that immediate NIV would have prevented reintubation in the original oxygen group.¹

The per-protocol analysis raises a separate issue. Excluding patients who did not meet baseline eligibility is conceptually different from excluding patients because they subsequently required an operation and reintubation. Removing procedural reintubations focuses attention on respiratory failure, but conditions the analysis partly on events after allocation and changes the outcome being examined. More procedural

reintubations were removed from the NIV group than the oxygen group. Its nominally significant result is informative as a sensitivity analysis, not a substitute for the primary randomised accounting.^{1,2}

The treatment delivered was a modest intermittent regimen. Median exposure of five hours on day one broadly met the minimum four-hour recommendation; median PEEP of 6 cmH₂O was within the permitted range but below the preferred 10. The trial therefore supports appraisal of this flexible, tolerance-adjusted package. It does not establish the optimal pressure or that longer continuous NIV would improve outcomes. Nor can one assume that a more intensive regimen would preserve the same tolerance profile. Mask fitting, staff attention, physiotherapy and prompt recognition of failure are part of implementing the comparison in practice.^{1,2}

The nested oxygen allocation is valuable because it reduces confounding by device choice. However, absence of an interaction does not demonstrate equivalent effects with HFNO and standard oxygen, and approximately 245 patients in each combination provide limited precision for patient-centred comparisons. The supplementary four-group outcomes were presented descriptively without formal pairwise testing. The data do not justify a universal ranking of all four combinations or a claim that HFNO must never replace NIV in every patient with obesity.^{1,2}

Selection also matters. Patients receiving home positive-pressure support or with recognised hypercapnia were excluded, removing groups with a particularly strong physiological reason for NIV. Conversely, many included surgical patients had been ventilated briefly for procedures, and a spontaneous breathing trial was not mandatory. The mixture broadens relevance but combines different baseline risks and reasons for needing support. Lack of subgroup interaction is not proof that these clinical settings are interchangeable. The evidence applies most directly to patients with obesity for whom clinicians had genuine uncertainty about preventive support, rather than those already requiring NIV for another indication.¹

The safety statement is correspondingly limited. No intervention-associated death or cardiac arrest is reassuring, but this narrow definition cannot exclude less catastrophic harms or treatment burden. Premature discontinuation was more common with NIV, and the trial did not demonstrate shorter ICU or hospital admission. The practical judgement is to consider preventive NIV when its likely value and tolerability justify the additional treatment, while recognising that EXTUB-OBESE alone establishes a reduction in strategy failure, not a confirmed improvement in survival or freedom from reintubation.¹

Body of Evidence

Prevention and rescue need to remain separate when interpreting postextubation NIV. In 2004, Esteban and colleagues randomised 221 patients who had already developed respiratory failure after elective extubation to mask NIV or standard therapy. Reintubation occurred in 48% of each group. ICU mortality was higher with NIV, 25% versus 14%, with a relative risk of 1.78 (95% confidence interval 1.03–3.20), and median time to reintubation was longer. The trial stopped after an interim analysis. It does not answer the preventive obesity question, but it cautions against treating rescue NIV as a harmless default or assuming that avoiding immediate intubation necessarily benefits the patient.⁶

El-Solh and colleagues then examined immediate preventive NIV in severe obesity. Sixty-two consecutive patients with BMI at least 35 kg/m² received NIV during the first 48 hours after extubation and were compared with 62 historically matched controls receiving conventional treatment. The absolute reduction in early respiratory failure was 16 percentage points (95% confidence interval 2.9–29.3). This was encouraging but non-randomised: historical matching could not exclude changes in selection or accompanying care. It supported a trial hypothesis without establishing the effect of routinely assigning NIV. Its more severely obese population and inclusion of hypercapnic patients also differ from EXTUB-OBESE.³

HIGH-WEAN supplied stronger randomised evidence in patients selected for high extubation risk because of age or chronic cardiac or respiratory disease. Of 648 randomised patients, 641 completed the trial. Immediate NIV alternating with HFNO reduced seven-day reintubation from 18.2% to 11.8%, an absolute difference of –6.4 percentage points (95% confidence interval –12.0 to –0.9). ICU mortality did not clearly differ. Treatment included at least 12 hours of NIV daily during the first 48 hours, substantially more than the minimum exposure in EXTUB-OBESE. This was a direct patient-centred benefit in a selected high-risk population, but obesity alone was not its enrolment criterion.⁴

A 2022 post hoc analysis of HIGH-WEAN examined 623 patients with BMI information. Among the 410 with overweight or obesity, reintubation occurred in 15 of 204 patients receiving NIV plus HFNO versus 41 of 206 receiving HFNO alone: difference –13 percentage points (95% confidence interval –19 to –6). There was evidence of heterogeneity by BMI, but the analysis was not prespecified. It combined overweight with obesity and reused participants from the parent trial; it is not a second independent replication. The findings strengthened the rationale for prospective assessment while leaving uncertainty about the most responsive population.⁵

EXTUB-OBESE adds that prospective obesity-specific comparison, with broader eligibility and both modern and conventional oxygen delivery. Its weaker reintubation signal need not contradict HIGH-WEAN. The populations differed, preventive NIV exposure was shorter, and the primary endpoint emphasised treatment switching over a three-day window rather than reintubation at seven days. These differences offer possible explanations, not proven causes of the different results. The randomised effect in one trial should not be strengthened retrospectively by borrowing favourable non-randomised analyses from another.¹

By the end of 2023, preventive NIV had credible support in selected patients at high risk, including encouraging evidence among those with increased BMI. EXTUB-OBESE showed that starting with NIV reduces subsequent failure of the initial support strategy in a broader obesity population. The remaining question was whether a defined preventive regimen produces enough reduction in reintubation or other meaningful outcomes to justify treating all such patients, including those who would recover with oxygen alone. Its results favour a considered, patient-specific use of NIV while preserving uncertainty about a universal policy.¹

Should we routinely use NIV immediately after extubation in all critically ill patients with obesity?

Consider it in selected patients. Treatment failure was reduced, but a patient-centred benefit from universal use remains unproven.

References

1. De Jong A, Bignon A, Stephan F, Godet T, Constantin JM, Asehnoune K, et al.; EXTUB-OBESE trial group. Effect of non-invasive ventilation after extubation in critically ill patients with obesity in France: a multicentre, unblinded, pragmatic randomised clinical trial. *Lancet Respir Med*. 2023;11(6):530–539.
2. De Jong A, Bignon A, Stephan F, Godet T, Constantin JM, Asehnoune K, et al. Supplementary appendix to: Effect of non-invasive ventilation after extubation in critically ill patients with obesity in France: a multicentre, unblinded, pragmatic randomised clinical trial. *Lancet Respir Med*. 2023.
3. El-Solh AA, Aquilina A, Pineda L, Dhanvantri V, Grant B, Bouquin P. Noninvasive ventilation for prevention of post-extubation respiratory failure in obese patients. *Eur Respir J*. 2006;28(3):588–595.

4. Thille AW, Muller G, Gacouin A, Coudroy R, Decavèle M, Sonnevile R, et al.; HIGH-WEAN Study Group and the REVA Research Network. Effect of Postextubation High-Flow Nasal Oxygen With Noninvasive Ventilation vs High-Flow Nasal Oxygen Alone on Reintubation Among Patients at High Risk of Extubation Failure: A Randomized Clinical Trial. *JAMA*. 2019;322(15):1465–1475.
5. Thille AW, Coudroy R, Nay MA, Gacouin A, Decavèle M, Sonnevile R, et al.; HIGH-WEAN Study Group and the REVA Research Network. Beneficial Effects of Noninvasive Ventilation after Extubation in Obese or Overweight Patients: A Post Hoc Analysis of a Randomized Clinical Trial. *Am J Respir Crit Care Med*. 2022;205(4):440–449.
6. Esteban A, Frutos-Vivar F, Ferguson ND, Arabi Y, Apezteguía C, González M, et al. Noninvasive Positive-Pressure Ventilation for Respiratory Failure after Extubation. *N Engl J Med*. 2004;350(24):2452–2460.

Possible Reporting Inconsistencies. The primary treatment-failure estimate labelled a relative risk, 0.43, appears to be an odds ratio: the Methods specify logistic regression, and the displayed counts, 66/490 versus 130/491, give a calculated crude odds ratio of 0.432 but a crude risk ratio of 0.509. The distinction affects the magnitude of the relative effect.^{1,2}

CAPE COD

**Dequin PF, Meziani F, Quenot JP, Kamel T, Ricard JD, Badie J, et al;
CRICS-TriGGERSep Network. Hydrocortisone in Severe Community-Acquired
Pneumonia. N Engl J Med. 2023;388(21):1931–1941.**

Introduction

Antibiotics treat the pathogen in community-acquired pneumonia, but do not immediately reverse the host response responsible for much of the associated organ dysfunction. In severe disease, pulmonary inflammation impairs gas exchange and can accompany circulatory failure and other systemic complications. An adjunct that limits harmful inflammation without compromising control of infection could therefore improve outcomes beyond those achieved by antimicrobials and supportive care alone.¹

Corticosteroids offer a plausible intervention, with important competing effects. Suppression of inflammation may reduce lung injury and progression to organ support, while hyperglycaemia, impaired immune responses and other adverse effects could offset that benefit. The balance is unlikely to be identical across non-severe pneumonia, severe hypoxaemic disease, influenza and established septic shock. Drug choice, timing and duration may also matter, but physiological plausibility does not establish which regimen produces a net clinical benefit.¹

Before CAPE COD, randomised evidence suggested improvements in clinical stability or composite treatment failure more consistently than improved survival. A small hydrocortisone study provided an encouraging early signal, while subsequent prednisone and methylprednisolone trials examined different populations and endpoints. The uncertainty was therefore not simply whether corticosteroids alter the course of pneumonia, but whether early administration can prevent death in patients requiring intensive care.^{2,3,4}

CAPE COD focused on that question using continuous intravenous hydrocortisone, begun soon after a qualifying severity criterion and adapted to clinical improvement. It deliberately excluded patients with recognised influenza and patients receiving vasopressors for septic shock at enrolment. This positioned the intervention as early treatment of severe community-acquired pneumonia rather than treatment of established septic shock. The primary hypothesis was a reduction in all-cause mortality at 28 days, providing a direct patient-centred test of the strategy.^{1,5}

Synopsis

CAPE COD was a phase 3, double-blind, placebo-controlled superiority trial conducted in 31 French centres. Recruitment occurred from 28 October 2015 to 11 March 2020. The French Ministry of Health funded the trial; neither the funder nor the coordinating sponsor, Tours University Hospital, participated in design, execution, interpretation or writing. Written consent came from patients or legally authorised representatives, with emergency inclusion and deferred consent permitted when neither could initially provide it.^{1,5}

Adults admitted to intensive or intermediate care required a clinical and radiological diagnosis of community-acquired pneumonia within 48 hours of hospital admission and at least one severity criterion. These were invasive or non-invasive ventilation with positive end-expiratory pressure at least 5 cmH₂O; high-flow nasal oxygen (HFNO) with inspired oxygen at least 50% and PaO₂/FiO₂ below 300 mmHg; qualifying hypoxaemia on a reservoir oxygen mask using a prespecified estimation chart; or Pneumonia Severity Index above 130. At least one antibiotic dose had to precede inclusion, and study treatment had to start within 24 hours of the first severity criterion. In April 2017, after 285 enrolments, the HFNO oxygenation threshold was widened from below 200 to below 300 and the mask criterion was added.^{1,5}

Exclusions included suspected aspiration, known influenza, recent invasive ventilation, prolonged antibiotic treatment for the current respiratory illness, post-obstructive pneumonia, cystic fibrosis, specified active infections, myelosuppression, treatment limitations and an existing indication for systemic corticosteroids. Prolonged prednisone above 15 mg/day or equivalent also precluded enrolment. Vasopressors for septic shock were excluded, but limited support attributed to sedation and positive-pressure ventilation was permitted if lactate was below 4 mmol/L and adrenaline or noradrenaline below 0.25 micrograms/kg/min. Subsequent septic shock could be treated with vasopressors at any dose.^{1,5}

A central web system allocated hydrocortisone or placebo equally using computer-generated blocks of four, stratified by centre and baseline mechanical ventilation. A statistician uninvolved in enrolment generated the sequence. Identical drug packages supported blinding. Both groups received antibiotics and clinician-directed respiratory and other supportive care; placebo was saline administered according to the same infusion and taper schedule. An unavoidable indication for open-label corticosteroids required stopping the study infusion, rather than adding the two treatments together.^{1,5}

Hydrocortisone was infused continuously at 200 mg/day for four days. Patients meeting all improvement criteria on day four—spontaneous breathing, $\text{PaO}_2/\text{FiO}_2$ above 200, no increase in Sequential Organ Failure Assessment (SOFA) score from baseline and a high probability of ICU discharge by day 14—then received 100 mg/day for two days and 50 mg/day for two days. Otherwise, the detailed supplementary regimen continued 200 mg/day through day seven, followed by 100 mg/day for four days and 50 mg/day for three days. Treatment stopped on ICU discharge, even before completing the nominal eight- or 14-day course.^{1,5}

The primary outcome was all-cause death by day 28. Secondary outcomes included day-90 mortality, ICU discharge, new ventilatory or vasopressor support, support-free days, oxygenation, organ dysfunction and day-90 quality of life. Support-free days began at randomisation, counted intervals between episodes and assigned zero to patients dying before day 28. Safety assessments included hospital-acquired infection, gastrointestinal bleeding, insulin requirements and weight change. Infection diagnoses were made by treating clinicians without independent adjudication.^{1,5}

The initial calculation sought a 25% relative mortality reduction, from 27.0% to 20.25%, with 80% power. Allowing two interim analyses and withdrawals increased the planned sample to 1200. The interim significance threshold was .001, with .049 reserved for the final analysis. The first review, in July 2018, included 398 patients and recommended continuation. Recruitment was suspended at 800 during the pandemic, while a separate embedded COVID-19 trial recruited a non-overlapping population. At the delayed second review in July 2021, $P=.0055$ did not meet the interim boundary. The monitoring board nevertheless recommended permanent cessation, citing the apparent benefit, ethical concerns and difficulty restarting recruitment.^{1,5}

The primary analysis used a chi-squared test and a Wald confidence interval for the absolute difference. The methods specified counting missing mortality outcomes as deaths, although the published day-90 table used reduced denominators. Fine-Gray models assessed ICU discharge with competing death, and new support or selected harms with death and ICU discharge as competing events. Secondary analyses were not adjusted for multiplicity. Mechanical-ventilation and pathogen-identification subgroups were prespecified; age, sex, severity-score and C-reactive-protein subgroups were post hoc.^{1,5}

Of 5948 patients screened, 800 were randomised, 401 to hydrocortisone and 399 to placebo. One hydrocortisone patient died before receiving treatment and was excluded; two placebo patients withdrew consent and two lacked legal capacity to consent. The primary population therefore comprised 400 and 395. Median age was 67 years and

approximately 69% were men. About 44% received mechanical ventilation, including approximately 22% invasively; 42% received HFNO. Median overall PaO₂/FiO₂ was 143 versus 137 mmHg. Only 54 patients qualified solely through the severity score. A pathogen was identified in approximately 55%, and most patients received a third-generation cephalosporin and a macrolide.^{1,5}

Study infusion began a median 15.3 versus 14.6 hours after ICU admission. Actual treatment lasted a median five versus six days. Discharge alive accounted for most early cessation: 261 versus 220 patients. Open-label corticosteroids prompted cessation in 17 versus 23, and influenza diagnosed after randomisation in 12 per group. Thus, the delivered comparison was an early, often relatively short course adjusted to recovery and diagnostic information.^{1,5}

By day 28, 25 of 400 patients (6.2%) had died with hydrocortisone versus 47 of 395 (11.9%) with placebo: absolute difference -5.6 percentage points (95% confidence interval -9.6 to -1.7; P=.006). The day-90 table reported 36 of 388 deaths (9.3%) versus 57 of 389 (14.7%); difference -5.4 points (-9.9 to -0.8). ICU discharge was more frequent with hydrocortisone, with a reported competing-risk hazard ratio of 1.33 (1.16–1.52).¹

Among patients receiving no mechanical ventilation at baseline, intubation occurred in 40 of 222 versus 65 of 220, with a reported hazard ratio of 0.59 (95% confidence interval 0.40–0.86). Among all initially non-intubated patients, including those receiving non-invasive ventilation, the corresponding figures were 60 of 308 versus 86 of 310, hazard ratio 0.69 (0.50–0.94). New vasopressors were used in 55 of 359 versus 86 of 344, hazard ratio 0.59 (0.43–0.82). Median ventilator- and vasopressor-free days were 28 in both groups, with different underlying distributions.^{1,5}

Serious adverse events affected 63 versus 88 patients, comprising 70 versus 99 events. Hospital-acquired infection occurred in 39 versus 44 patients and gastrointestinal bleeding in nine versus 13, with wide intervals around their competing-risk estimates. Insulin was administered to 231 versus 177 patients. Among insulin recipients, median daily dose during the first week was 35.5 versus 20.5 units. Day-90 quality-of-life component scores were similar among the 201 versus 174 respondents.^{1,5}

Critique

CAPE COD provides persuasive evidence of benefit from early hydrocortisone in its selected population, but the strength and size of that benefit require qualification. Its primary endpoint is death, rather than a surrogate or clinician-defined composite, and the direction of several support-related outcomes is coherent with the mortality finding.

The most important concern is how recruitment ended: the trial stopped below its planned interim efficacy threshold, after a prolonged pandemic interruption. That distinction belongs alongside the result, not hidden behind the phrase “stopped after an interim analysis”.^{1,5}

The monitoring board had reasons beyond a formal statistical boundary, including concern about withholding an apparently beneficial treatment and practical barriers to restarting. However, the second interim P value of .0055 did not meet the prespecified .001 criterion. The reported final threshold of .049 cannot retrospectively make that an interim boundary crossing. This discretionary stopping decision leaves greater uncertainty about the operating characteristics and potential exaggeration of benefit than a fully completed or strictly boundary-stopped trial. It does not erase the observed mortality difference, but weakens a claim that the planned sequential design was satisfied without qualification.^{1,5}

Precision is further affected by the unexpectedly low control mortality. The study anticipated 27%, yet observed 11.9%, and analysed approximately two-thirds of its target population. There were only 72 primary deaths. The absolute interval still excludes no difference in the reported analysis, but spans benefits of substantially different practical magnitude. Calculated directly from the primary proportions, the absolute reduction is approximately 5.65 percentage points, equivalent to about 18 patients treated per additional survivor at day 28. This descriptive estimate should not be assumed to remain constant in populations with different risks or after further replication.¹

The selection criteria help explain what was tested. Severe hypoxaemia was common, but patients with vasopressor-treated septic shock were excluded. The low mortality does not mean the cohort had trivial disease: many required non-invasive or invasive ventilation and oxygenation was markedly impaired. Rather, this was a population with significant respiratory illness from which several particularly high-risk or clinically distinct groups had been removed. The trial should not be used to settle treatment in septic shock, aspiration, recognised influenza or profound immunosuppression.^{1,5}

The allowance for vasopressors after sedation or positive-pressure ventilation is a useful nuance. Approximately 12% received baseline vasopressors despite the septic-shock exclusion, under restricted dose and lactate criteria. Calling the population entirely vasopressor-free would be inaccurate. Conversely, these patients do not make the trial representative of unrestricted septic shock. The criteria depended partly on clinicians attributing circulatory impairment to treatment rather than infection, which introduces

some classification uncertainty without demonstrating that allocation itself was biased.^{1,5}

Timing is another defining feature. Treatment began within 24 hours of a qualifying severity criterion and approximately 15 hours after ICU admission. This makes the trial a test of early intervention, often before invasive ventilation. The larger numerical mortality separation in patients not receiving mechanical ventilation is compatible with prevention of deterioration, but does not establish that ventilation status identifies responders. Subgroup estimates are imprecise, and several additional subgroup definitions were post hoc. Similarly, a C-reactive-protein threshold should not be converted into a validated selection rule from exploratory findings.^{1,5}

Central concealed randomisation, placebo control and blinding strengthen causal interpretation, particularly for decisions about support and discharge. Glucocorticoid effects such as altered glucose requirements might occasionally suggest allocation, but the study does not establish that blinding failed. The primary mortality endpoint is comparatively resistant to subjective ascertainment. There is nevertheless a deviation from full intention-to-treat accounting: the patient who died after hydrocortisone allocation but before the first dose was omitted. Including that known death would change the treated mortality proportion slightly, not remove the overall direction, but exclusion on this basis is methodologically undesirable.¹

Longer-term mortality is supportive but less securely reported. The methods describe assigning death to missing mortality values, whereas Table 2 presents day-90 outcomes for 388 and 389 patients rather than 400 and 395. The supplied results do not reconcile that difference. Accordingly, the chapter retains the published denominators and does not portray day-90 survival as fully ascertained for every participant. This reporting uncertainty is material to the durability of benefit, although it does not prevent appraisal of the clearly presented primary comparison.¹

Treatment duration should also be interpreted as part of a strategy. Most patients stopped before the nominal eight- or 14-day endpoint, commonly because they left intensive care. That was intended behaviour, not wholesale non-adherence. The randomised result therefore cannot justify automatically completing 14 days after ICU discharge. Nor did the trial compare continuous infusion with intermittent dosing, tapering with abrupt cessation, or hydrocortisone with another corticosteroid. Differences in pharmacology and timing may explain disagreements between trials, but CAPE COD did not randomise those explanations.^{1,5}

The support-related findings are clinically encouraging, with appropriate attention to death and discharge as competing events. The subgroup at risk of intubation must be specified: those receiving no ventilation and those receiving no invasive ventilation are overlapping populations, not independent confirmations. Secondary confidence intervals were not adjusted for multiplicity. Similarly, identical median support-free days do not negate the mortality result; more than half the patients could accumulate 28 days, producing a ceiling that conceals other distributional differences. Survivor-only length of stay and incomplete quality-of-life responses cannot establish identical recovery among all randomised patients.^{1,5}

Safety assessment supports cautious use rather than a declaration of no harm. Infection and bleeding counts did not show a clear excess, but their intervals permit clinically relevant differences. Infection was clinician-reported, and earlier ICU discharge changed the opportunity to observe ICU-acquired complications. Insulin use and requirements increased, while the reversibility of hyperglycaemia was not established. Neuropsychiatric and neuromuscular complications were not specifically assessed. These limits favour close metabolic monitoring and attention to patient selection. On the evidence available in 2023, early hydrocortisone is a reasonable treatment for patients resembling the trial population, with the stopping decision and unresolved long-term safety limiting certainty about the exact benefit.^{1,5}

Body of Evidence

The early hydrocortisone evidence came from a 46-patient, multicentre, placebo-controlled trial by Confalonieri and colleagues. A 200-mg intravenous bolus followed by 10 mg/hour for seven days improved oxygenation, organ-dysfunction measures and delayed septic shock. The primary objectives emphasised physiological recovery and complications; the abstract also described lower mortality, but the very small sample limited confidence in its size. This study supplied a rationale for larger investigation of hydrocortisone, rather than a definitive survival estimate or validation of CAPE COD's different dose and adaptive duration.²

The 2015 prednisone trial addressed a broader hospitalised population. Among 785 patients, 50 mg daily for seven days shortened median time to clinical stability from 4.4 to 3.0 days, hazard ratio 1.33 (95% confidence interval 1.15–1.50). Hyperglycaemia requiring insulin occurred more often, 19% versus 11%. The primary endpoint captured an earlier stage of recovery than death and did not establish a mortality benefit in severe ICU pneumonia. Its main contribution was evidence that corticosteroids can accelerate clinical improvement while imposing a measurable metabolic cost.³

Torres and colleagues studied a more selected inflammatory phenotype: 120 patients with severe pneumonia and C-reactive protein above 150 mg/L. Methylprednisolone 0.5 mg/kg every 12 hours for five days, started within 36 hours of hospital admission, reduced composite treatment failure from 31% to 13%, odds ratio 0.34 (95% confidence interval 0.14–0.87). The composite included radiographic progression as well as shock, ventilation and death. In-hospital mortality did not clearly differ. CAPE COD advances beyond this surrogate-heavy result by using mortality as the primary endpoint, but did not require the same inflammatory threshold.⁴

ESCAPE, published in 2022, provided the closest large severe-pneumonia comparison. It randomised 586 patients at 42 US Veterans Affairs centres, analysing 584 after recruitment fell well short of the 1420 target. Methylprednisolone was given as a 40-mg loading dose followed by 40 mg/day through day seven and tapering over 20 days; recruitment occurred within 72–96 hours of hospital presentation. Sixty-day mortality was 16% versus 18%, adjusted odds ratio 0.90 (95% confidence interval 0.57–1.40). This did not establish benefit, but its interval and recruitment shortfall also preclude treating it as definitive evidence of no effect. Compared with CAPE COD, drug, treatment window, duration and population differed; none was independently randomised as the cause of the contrasting findings.⁶

A later contribution within the book's cut-off came from the December 2023 follow-up of the prednisone trial. This was a companion analysis, not a new randomised population. Among 727 patients available at 180 days, mortality did not clearly differ, but recurrent pneumonia and secondary infections were more frequent after prednisone, with odds ratios of 2.57 (95% confidence interval 1.29–5.12) and 1.94 (1.25–3.03). New insulin dependence was also more frequent, with considerable uncertainty. Different severity and treatment prevent direct transfer of those estimates to CAPE COD, but they reinforce why short-term ICU safety observations cannot settle the longer-term balance.⁷

Taken together, the evidence available by the end of 2023 moved the question beyond whether inflammation can be suppressed. CAPE COD demonstrated a mortality signal with early hydrocortisone in a defined severe-pneumonia population, supported by less subsequent organ support. Earlier trials showed that a beneficial physiological or clinical-stability effect does not guarantee survival benefit, and the follow-up evidence maintained concern about delayed harms. The most defensible application is therefore to reproduce the relevant patient selection, early timing and recovery-adjusted course, while avoiding claims that every corticosteroid regimen or every pneumonia population shares the same benefit.¹

Should we give early hydrocortisone for severe community-acquired pneumonia in patients resembling CAPE COD participants?

Yes, it is reasonable in appropriately selected patients. The mortality benefit is encouraging, with uncertainty from early stopping and limited longer-term safety assessment.

References

1. Dequin PF, Meziani F, Quenot JP, Kamel T, Ricard JD, Badie J, et al.; CRICS-TriGGERSep Network. Hydrocortisone in Severe Community-Acquired Pneumonia. *N Engl J Med*. 2023;388(21):1931–1941.
2. Confalonieri M, Urbino R, Potena A, Piattella M, Parigi P, Puccio G, et al. Hydrocortisone Infusion for Severe Community-acquired Pneumonia: A Preliminary Randomized Study. *Am J Respir Crit Care Med*. 2005;171(3):242–248.
3. Blum CA, Nigro N, Briel M, Schuetz P, Ullmer E, Suter-Widmer I, et al. Adjunct prednisone therapy for patients with community-acquired pneumonia: a multicentre, double-blind, randomised, placebo-controlled trial. *Lancet*. 2015;385(9977):1511–1518.
4. Torres A, Sibila O, Ferrer M, Polverino E, Menendez R, Mensa J, et al. Effect of Corticosteroids on Treatment Failure Among Hospitalized Patients With Severe Community-Acquired Pneumonia and High Inflammatory Response: A Randomized Clinical Trial. *JAMA*. 2015;313(7):677–686.
5. Dequin PF, Meziani F, Quenot JP, Kamel T, Ricard JD, Badie J, et al. Supplementary appendix to: Hydrocortisone in Severe Community-Acquired Pneumonia. *N Engl J Med*. 2023.
6. Meduri GU, Shih MC, Bridges L, Martin TJ, El-Solh A, Seam N, et al.; ESCAPe Study Group. Low-dose methylprednisolone treatment in critically ill patients with severe community-acquired pneumonia. *Intensive Care Med*. 2022;48(8):1009–1023.
7. Blum CA, Roethlisberger EA, Cesana-Nigro N, Winzeler B, Rodondi N, Blum MR, et al. Adjunct prednisone in community-acquired pneumonia: 180-day outcome of a multicentre, double-blind, randomized, placebo-controlled trial. *BMC Pulm Med*. 2023;23(1):500.

ICONIC

van der Wal LI, Grim CCA, del Prado MR, van Westerloo DJ, Boerma EC, Rijnhart-de Jong HG, et al.; ICONIC investigators. Conservative versus Liberal Oxygenation Targets in Intensive Care Unit Patients (ICONIC): A Randomized Clinical Trial. Am J Respir Crit Care Med 2023;208:770–779.

Introduction

Oxygen administration poses an unusually familiar therapeutic uncertainty. Correcting inadequate oxygenation is essential, but the appropriate upper target once immediate hypoxaemia has resolved is less clear. Excess oxygen can promote absorption atelectasis, pulmonary inflammation and vasoconstriction. Conversely, reducing oxygenation may diminish the margin for an unexpected deterioration in gas exchange or perfusion. These competing mechanisms make it difficult to select a universal target from physiology alone. The relevant question is whether deliberately maintaining different oxygenation ranges improves outcomes in the patients actually exposed to them.¹

An early signal favouring conservative oxygen therapy came from Oxygen-ICU, a single-centre trial published in 2016. Its mortality findings encouraged attention to unnecessary oxygen exposure, but early termination and the limited population required confirmation. ICONIC began recruitment in 2018 against this background. Subsequent trials became available during its conduct and before publication, making it necessary to distinguish the rationale for initiating the study from the evidence against which its final results should be judged.^{1,2}

An important difficulty in oxygen research is that assigned targets and achieved exposure may differ substantially. A nominally liberal group may spend little time with appreciable hyperoxaemia, while a conservative group may remain above its intended range without supplemental oxygen. Small physiological differences can then make a neutral clinical result difficult to interpret. ICONIC deliberately compared widely separated arterial oxygen targets, with treatment beginning very early in invasive ventilation and continuing after extubation while patients remained in intensive care.¹

The trial therefore offered an opportunity to test whether a meaningful oxygenation contrast would alter survival. Its eventual contribution depends on two distinct questions: whether the intervention changed oxygen exposure, and whether the available randomised comparison was sufficiently complete and precise to identify a

clinically important benefit or harm. Successful separation alone cannot answer the second question.

Synopsis

ICONIC was an investigator-initiated, open-label, parallel-group superiority trial in eight Dutch ICUs and one Italian ICU. The Dutch Research Council funded the study without involvement in its design, analysis or publication decision. Adults requiring intubation and mechanical ventilation for an anticipated duration of at least 24 hours were eligible. Inclusion was permitted within two hours of starting invasive ventilation in the ICU, or within two hours of ICU admission for patients already intubated. Ventilation for longer than 12 hours immediately before admission was an exclusion.^{1,3}

Important exclusions were a decision to withhold life-sustaining treatment, acute respiratory distress syndrome (ARDS) with a $\text{PaO}_2/\text{FiO}_2$ ratio below 150 mmHg, acute decompensation of chronic obstructive pulmonary disease (COPD) and chronic hypoxaemia, home oxygen therapy, severe pulmonary hypertension, and severe low cardiac output shock not rapidly reversible. Other exclusions included venoarterial extracorporeal membrane oxygenation, a specific indication for hyperoxygenation, severe anaemia not rapidly reversible, uncontrollable intracranial hypertension, pregnancy and relevant competing interventional studies. These restrictions selected patients in whom both target ranges were considered practicable.^{1,3}

Secure web-based allocation used computer-generated variable blocks, a 1:1 ratio and stratification by site. Clinicians and outcome assessors knew the allocation; data analysts remained blinded. Deferred consent was permitted. Patients who died before delayed consent could be obtained remained in the analysis, but those without consent within five days were excluded. Exclusion criteria discovered after randomisation could also lead to withdrawal when present at enrolment.¹

The lower target was PaO_2 55–80 mmHg, with a pulse oximetry oxygen saturation (SpO_2) target of 91–94%. The higher target was PaO_2 110–150 mmHg or SpO_2 96–100%. Treatment continued until ICU discharge or day 28, whichever occurred first. During mechanical ventilation, at least one arterial blood sample per shift was required, corresponding to three daily measurements. Clinicians could adjust the fraction of inspired oxygen (FiO_2) and positive end-expiratory pressure (PEEP), guided by a recommended combination table that allowed individual departures.^{1,3}

There were safeguards against excessive ventilatory treatment solely to attain the higher target. If FiO_2 exceeded 0.8 or PEEP exceeded 14 cmH₂O for more than two

hours, clinicians could temporarily reduce FiO_2 to 0.8 and limit PEEP to 15 cmH₂O, reassessing target feasibility every two hours. Proning, recruitment manoeuvres and extracorporeal support required clinical indications. Brief high oxygen exposure during airway procedures was allowed. After extubation, supplemental oxygen was generally avoided in the lower group unless PaO_2 fell below 55 mmHg; the higher group received nasal oxygen at 5 L/min unless PaO_2 exceeded 150 mmHg.¹

The primary outcome was all-cause mortality at 28 days. Secondary outcomes included ICU, hospital and 90-day mortality, ICU and hospital length of stay, ventilator-free days at day 28 and ischaemic events. Ventilator-free days counted time alive without invasive ventilation, requiring at least 24 consecutive hours of unassisted breathing. Specified serious adverse events included PaO_2 below 37.5 mmHg, SpO_2 below 80% for more than ten minutes, cardiac arrest, and intestinal, cerebral, cardiac or limb ischaemia.¹

The planned sample of 1,512 assumed 24% control mortality and sought a six percentage point absolute reduction with 80% power and a two-sided significance level of 0.05. The principal analysis was modified intention to treat, excluding patients without consent or subsequently recognised as ineligible. Additional analyses accounted for site and examined patients achieving the assigned oxygenation thresholds. Mortality monitoring was added after the study began; an interim analysis after 500 patients found no significant in-hospital mortality difference. Recruitment ultimately stopped because pandemic workload and slow enrolment were expected to require another five years to complete the intended sample.¹

Between November 2018 and November 2021, 972 patients were screened and 882 randomised: 439 to lower and 443 to higher oxygenation. Subsequent exclusions removed 104 and 114, leaving 335 and 329 in the primary analysis. Consent refusal, withdrawal or inability to obtain consent accounted for 93 and 88 exclusions; recognised exclusion criteria accounted for another 11 and 26. There was no loss to outcome follow-up among the 664 analysed patients. Median age was 67 years in both groups and median admission Sequential Organ Failure Assessment (SOFA) score was nine. Approximately three quarters were medical admissions; cardiac arrest accounted for 26.6% and 29.2%.¹

During mechanical ventilation, median achieved PaO_2 was 75 versus 115 mmHg, and SpO_2 95% versus 99%, in lower versus higher groups. Median FiO_2 was 35.5% versus 48.1%, with PEEP of 8 versus 9 cmH₂O. Median invasive ventilation duration was 3.0 versus 2.8 days. After mechanical ventilation, median PaO_2 was 75 versus 85.5 mmHg, indicating a smaller arterial oxygen contrast despite continuing separation in SpO_2 .^{1,3}

At day 28, 129/335 patients (38.5%) in the lower group and 114/329 (34.7%) in the higher group had died. The reported risk ratio was 1.11 (95% confidence interval 0.9–1.4). The abstract and accompanying relative-risk statement gave $P=0.30$; the principal outcome table and supplementary analysis gave $P=0.34$. Both analyses failed to demonstrate superiority. At day 90, mortality was 144/335 (43.0%) versus 133/329 (40.4%; $P=0.56$). Median ventilator-free days were 18.3 versus 20.2 ($P=0.36$); ICU and hospital lengths of stay also showed no clear difference.^{1,3}

Thirteen versus 22 serious adverse events occurred in 12 versus 17 patients. Ischaemic events numbered ten versus 15, including four versus seven intestinal events. No PaO_2 below 37.5 mmHg was recorded as a serious adverse event. The analysis restricted by achieved oxygenation included 229 lower-target and 171 higher-target patients; day-28 mortality was 82/229 (35.8%) versus 67/171 (39.2%; $P=0.57$). Exploratory subgroup analyses did not establish a population with a reliable treatment benefit.^{1,3}

Critique

ICONIC demonstrated substantial physiological separation without establishing which strategy improved survival. These are compatible findings. The analysed sample reached only 43.9% of the intended size, and mortality was numerically higher with the lower target. The reported relative-risk interval includes both benefit and substantial harm. An independently calculated unadjusted risk difference is 3.9 percentage points, with an approximate 95% confidence interval from –3.5 to 11.2 percentage points. This interval is too wide to support equivalence or an assurance that clinically important effects have been excluded. The result is an imprecise superiority comparison.¹

Premature termination explains much of that imprecision. Pandemic disruption was an operational reason for stopping, rather than evidence that the treatments had become interchangeable. The original six percentage point treatment effect was also ambitious for a broadly applied supportive therapy. Smaller survival differences would still matter when an intervention is inexpensive and used so frequently. Control mortality of 34.7% exceeded the anticipated 24%, but a higher event rate does not restore the missing sample or make the confidence interval sufficiently narrow. The trial adds information; it does not close the oxygen target question.^{1,4}

The larger threat to causal interpretation is the removal of 218 patients after randomisation and treatment assignment. This represents 24.7% of those randomised, with exclusions of 23.7% and 25.7% in the two groups. Consent procedures enabled exceptionally early recruitment, and retaining patients who died before consent reduced one obvious mechanism of selective exclusion. Nevertheless, refusal, transfer and

inability to complete consent could relate to prognosis or treatment experience. Similar overall exclusion proportions cannot establish that the excluded populations were comparable. Their characteristics and outcomes were unavailable, so the direction and magnitude of any resulting selection bias cannot be determined.¹

The oxygenation contrast is a substantive strength. A 40 mmHg difference in median PaO₂ during ventilation makes inadequate arterial separation an unconvincing complete explanation for the result. Early enrolment also limited exposure to shared ICU oxygen management before allocation. However, the intervention should be interpreted as delivered: the lower group's median SpO₂ was 95%, above its assigned 91–94% range, and its median PaO₂ was near the upper boundary of 55–80 mmHg. ICONIC therefore does not demonstrate the safety of maintaining patients persistently at the lower boundary or tolerating prolonged saturations below the tested experience.^{1,3}

The magnitude and duration of exposure also need separating. Median invasive ventilation lasted approximately three days, although targets could continue for 28 days. After extubation, the median PaO₂ difference narrowed to 10.5 mmHg, and later daily plots represented progressively fewer patients. The aggregate PaO₂ contrast cannot be treated as a uniform 28-day difference for every participant. Blood gases were intermittent, and arterial data were unavailable for eight analysed patients. These limitations do not negate the achieved intervention, but restrict conclusions about cumulative oxygen dose, transient extremes and patients requiring prolonged ventilation.^{1,3}

Assignment changed a treatment strategy rather than PaO₂ in isolation. The higher group received more inspired oxygen and slightly higher PEEP, while rescue treatment and temporary relaxation of targets remained available. Any clinical effect would therefore combine arterial oxygenation, pulmonary oxygen exposure and the consequences of ventilator adjustment. Conversely, these safeguards made the comparison more clinically defensible. It would be inappropriate to extrapolate its findings to unrestricted FiO₂ escalation, prolonged extreme hyperoxaemia or aggressive recruitment undertaken simply to force a laboratory target.^{1,3}

The analysis restricted by achieved PaO₂ is particularly difficult to interpret. Eligibility depended on at least half of blood gases being at or below 80 mmHg in the lower group, or at or above 110 mmHg in the higher group. This can exclude lower-group patients with good gas exchange who exceed 80 mmHg on minimal oxygen, while excluding higher-group patients with poor gas exchange who cannot reach 110 mmHg despite substantial support. The supplement explicitly recognises this asymmetry. Only 400 patients remained, with unequal group retention and a reversed numerical mortality

difference. Selecting on postrandomisation physiology cannot resolve uncertainty in the randomised comparison.^{1,3}

Mortality was an appropriate, objective primary outcome, and complete follow-up within the analysed cohort is reassuring. Lack of blinding remains relevant to co-interventions and treatment withdrawal, even when death itself is unambiguous. Ventilator-free days and length of stay are additionally affected by clinical decisions and competing mortality. Site adjustment yielded similar probabilities, but supplementary reporting largely provides P-values rather than adjusted effect estimates and intervals. The $P=0.30$ versus $P=0.34$ discrepancy is numerically compatible with different chi-square implementations; the supplied report does not explicitly reconcile them. Its practical importance is small compared with selection and imprecision.^{1,3}

Safety reporting also warrants restraint. Twelve versus 17 patients experienced serious adverse events, and intestinal ischaemia was uncommon. These counts provide no clear harm signal favouring either strategy, but are insufficient to establish equal safety. Similarly, the numerous exploratory subgroup comparisons report within-subgroup probabilities rather than convincing treatment interaction evidence. A nonsignificant comparison in a small diagnostic group does not establish absence of heterogeneity. Neither the apparently favourable cardiac arrest direction nor the opposite stroke direction provides a dependable basis for assigning patient-specific targets.^{1,3}

Applicability is bounded by the eligibility restrictions. Patients with more severe oxygenation impairment from ARDS, acute COPD decompensation, chronic hypoxaemia and several conditions limiting oxygen delivery were excluded. Stable COPD was nevertheless present in both groups, so describing the cohort as excluding all COPD would be inaccurate. The appreciable cardiac arrest population also makes this different from a trial confined to pulmonary respiratory failure. The defensible conclusion is that lower targeting did not demonstrate a survival advantage in this selected population; it does not identify an optimum for every ventilated patient.^{1,3}

Body of Evidence

Oxygen-ICU provided an influential early efficacy signal. Among 434 patients in its modified intention-to-treat analysis, ICU mortality was 11.6% with conservative oxygen and 20.2% with conventional oxygen, an absolute reduction of 8.6 percentage points (95% confidence interval 1.7–15.0). Achieved median PaO_2 values were 87 and 102 mmHg. The trial stopped before its planned sample was reached and came from one centre. Its large mortality difference was therefore a reason to pursue multicentre confirmation. Its conservative exposure was also appreciably higher than ICONIC's lower

target, illustrating why treatment labels alone do not define comparable oxygen strategies.²

ICU-ROX subsequently randomised 1,000 ventilated adults to conservative or usual oxygen therapy. Both groups used a default lower SpO₂ limit of 90%; the conservative strategy reduced oxygen exposure using an upper saturation alarm and downward FiO₂ titration. Median ventilator-free days were 21.3 versus 22.1, with a site-adjusted difference of -0.3 days (95% confidence interval -2.1 to 1.6; P=0.80). Mortality at 180 days was 35.7% versus 34.5%. This weakened the case for a large general benefit from conservative oxygen, but usual care was a different comparator from ICONIC's deliberately high PaO₂ target. The trials address related, rather than identical, treatment decisions.⁵

LOCO₂ tested a lower target in a specifically pulmonary population. Patients with ARDS received PaO₂ targets of 55–70 or 90–105 mmHg for seven days. The trial stopped after 205 enrolments because of safety concerns and a low likelihood of showing its intended primary benefit; 201 patients were analysed. Day-28 mortality was 34.3% versus 26.5%, an inconclusive difference, while day-90 mortality was 44.4% versus 30.4%, a difference of 14.0 percentage points (95% confidence interval 0.7–27.2). Five mesenteric ischaemic events occurred in the conservative group and none in the liberal group. These small-study findings require caution, but ICONIC cannot dismiss them: it excluded ARDS with PaO₂/FiO₂ below 150 mmHg and delivered a higher median oxygenation in its lower group.^{1,6}

HOT-ICU was a larger mortality trial in acute hypoxaemic respiratory failure. It randomised 2,928 patients to PaO₂ targets of 60 or 90 mmHg, with eligibility based on substantial supplemental oxygen requirements. At 90 days, 618/1,441 patients (42.9%) in the lower-target group and 613/1,447 (42.4%) in the higher-target group had died: adjusted risk ratio 1.02 (95% confidence interval 0.94–1.11; P=0.64). This more precise estimate supplied stronger evidence against a large mortality advantage from lower targeting in that population. Nevertheless, its higher target was below ICONIC's 110–150 mmHg range. The absence of benefit across these comparisons does not make all higher targets interchangeable.⁷

PILOT approached the question through three SpO₂ targets: 90%, 94% and 98%, with corresponding ranges of 88–92%, 92–96% and 96–100%. This pragmatic cluster-crossover trial at one academic centre included 2,541 patients in its primary analysis. Median ventilator-free days were 20, 21 and 21 (P=0.81); in-hospital death by day 28 occurred in 34.8%, 34.0% and 33.2%, respectively. Including an intermediate target helped address the concern that a comparison of two extremes might overlook a

better middle range. The neutral primary result did not establish a uniquely preferred saturation target or rule out smaller mortality differences.⁸

By the end of 2023, these studies had moved the discussion beyond the expectation that conservative oxygen would produce a large, consistent benefit throughout critical care. They had not established a single optimal range. ICONIC adds a conspicuous arterial oxygen contrast, but its incomplete analysis population and wide uncertainty limit the additional clinical certainty it provides. The accompanying editorial argued for much larger pragmatic comparisons capable of detecting modest effects and for validated approaches to treatment heterogeneity. That was a research direction, rather than evidence for an individualised prescribing rule already available at the book's cut-off. The practical lesson is to keep the tested population, achieved exposure and statistical precision visible whenever an oxygen trial is used to justify a target.^{1,4}

Does targeting PaO₂ 55–80 mmHg improve survival compared with 110–150 mmHg in ventilated ICU patients?

ICONIC did not demonstrate a survival benefit from the lower target. Early termination and substantial exclusions after randomisation prevent a conclusion that the strategies are equivalent.

References

1. van der Wal LI, Grim CCA, del Prado MR, van Westerloo DJ, Boerma EC, Rijnhart-de Jong HG, et al.; ICONIC investigators. Conservative versus Liberal Oxygenation Targets in Intensive Care Unit Patients (ICONIC): A Randomized Clinical Trial. *Am J Respir Crit Care Med* 2023;208:770–779.
2. Girardis M, Busani S, Damiani E, Donati A, Rinaldi L, Marudi A, et al. Effect of Conservative vs Conventional Oxygen Therapy on Mortality Among Patients in an Intensive Care Unit: The Oxygen-ICU Randomized Clinical Trial. *JAMA* 2016;316:1583–1589.
3. ICONIC investigators. Online data supplement to: Conservative versus Liberal Oxygenation Targets in Intensive Care Unit Patients (ICONIC): A Randomized Clinical Trial. *Am J Respir Crit Care Med* 2023;208:770–779.
4. DeMasi SC, Self WH, Semler MW. Oxygen Targets and the Future of Critical Care Clinical Research. *Am J Respir Crit Care Med* 2023;208:746–748.

5. 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;382:989–998.
6. 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;382:999–1008.
7. 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;384:1301–1311.
8. 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;387:1759–1769.



Gastrointestinal Trials

EIPAN

Jabaudon M, Genevrier A, Jaber S, Windisch O, Bulyez S, Laterre PF, et al.; EIPAN study group. Thoracic epidural analgesia in intensive care unit patients with acute pancreatitis: the EIPAN multicenter randomized controlled trial. Crit Care 2023;27:213.

Introduction

Acute pancreatitis presents an unusual opportunity for an analgesic intervention to become a disease-modifying treatment. Abdominal pain is an immediate therapeutic priority, while patients admitted to intensive care may also develop respiratory failure, circulatory instability and abdominal complications. Systemic opioids provide analgesia, but reducing their requirement could be useful if this can be achieved without compromising comfort or introducing greater procedural risk. The more ambitious question is whether the method of analgesia can alter the course of pancreatitis itself.¹

Thoracic epidural analgesia combines local anaesthetic administration with segmental sympathetic blockade. Improved pancreatic and splanchnic perfusion offers a plausible route from pain treatment to reduced tissue injury. Experimental work supported this mechanism, while a small clinical trial suggested improved pancreatic perfusion. Neither establishes that epidural treatment prevents respiratory failure or improves survival in patients with an evolving, heterogeneous inflammatory illness.^{2,3}

The balance is particularly uncertain in intensive care. Epidural insertion requires expertise, and coagulopathy, infection or refractory shock may prevent its use. Sympathetic blockade may be undesirable in a patient whose circulation already depends on vasopressors. Sedation and mechanical ventilation can make assessment of the sensory block difficult. A technically successful catheter therefore does not necessarily establish that the intended physiological intervention has been delivered.^{1,4}

EIPAN addressed both the clinical promise and this practical uncertainty. It compared thoracic epidural analgesia added to usual care with usual care alone in adults admitted to intensive care with acute pancreatitis. Its hypothesis was that the intervention would increase days alive without invasive ventilation during the first month. This moved the question beyond analgesic efficacy to whether an invasive analgesic strategy could improve a patient-centred consequence of organ dysfunction.¹

Synopsis

EPIPAN was a pragmatic, open-label, parallel-group superiority trial conducted in 11 intensive care units in France, Switzerland and Belgium. Recruitment ran from June 2014 to January 2019. Funding came from the French Society of Anaesthesia and Intensive Care Medicine and Clermont-Ferrand University Hospital; the sponsors had no role in study conduct, analysis or publication. The investigators declared no competing interests.¹

Adults were eligible with acute pancreatitis on intensive care admission, diagnosed by at least two of characteristic pain, lipase at least three times the upper normal limit, and compatible computed tomography. Important exclusions included prothrombin time below 60%, platelets below $75 \times 10^9/L$, insufficient interruption of therapeutic heparin, local or central nervous system infection, relevant previous spinal procedures, intracranial hypertension and refractory shock. Written consent was obtained from patients or their representatives.¹

A central web system allocated patients equally using concealed, variable blocks. Randomisation was stratified by centre, symptom duration below or at least 48 hours, and the highest respiratory, renal or haemodynamic Marshall score, grouped as zero, 1–2 or 3–4. Clinicians and patients were unblinded; final outcome assessment, laboratory measurements and statistical analysis were blinded.¹

The intervention was epidural analgesia plus usual care. A qualified anaesthetist-intensivist or supervised resident inserted a catheter at T6–T9 as soon as possible. Ropivacaine 2 mg/mL with sufentanil 0.5 µg/mL was infused at 5–15 mL/hour, with patient- or nurse-controlled boluses of 3–10 mL and a ten-minute lockout. Treatment was intended to last at least 72 hours; epidural clonidine was optional. Weaning and catheter removal followed local practice.¹

Both groups received multimodal analgesia, including systemic opioid and non-opioid drugs as required. The targets were a visual analogue pain score below 40/100 or a behavioural pain score of 3–4, assessed at least every four hours. Controls did not routinely receive an epidural. Fluid and nutritional management followed broad clinical recommendations, with substantial bedside discretion. Protective ventilation and weaning were encouraged, without a single detailed mandatory pathway governing every intubation or extubation decision.^{1,4}

The primary outcome was ventilator-free days from randomisation to day 30: days alive without invasive ventilatory assistance, with deaths before day 30 assigned zero. Brief ventilation lasting less than 24 hours for surgery did not reduce this outcome. Higher

values were favourable. Secondary outcomes included mortality, organ failure, abdominal complications, ventilation duration, pain, feeding tolerance and serial markers of inflammation, lung injury and kidney injury. Blood samples assessed interleukin-6, soluble receptor for advanced glycation end-products and neutrophil gelatinase-associated lipocalin at baseline and on days two and seven. Urinary tissue inhibitor of metalloproteinases-2 and insulin-like growth factor-binding protein-7 were also measured. Planned economic results were unavailable.^{1,4}

The sample-size calculation assumed a control mean of 13 ventilator-free days with a standard deviation of 15. A total of 148 participants provided 80% power to detect an increase of seven days at a two-sided 5% significance level. The principal comparison used the Mann–Whitney test. Additional modelling addressed the concentration of observations at 30 days, with adjusted analyses incorporating stratification factors. Subgroup analyses and this distribution-focused modelling were post hoc. An independent safety committee reviewed the planned interim analysis after 74 participants, including unblinded adverse-event information. No early termination was described.^{1,4}

Of 316 assessed patients, 148 were randomised, 74 per group. Refusal of consent and contraindications, especially coagulation abnormalities, accounted for substantial non-enrolment. The primary analysis contained 65 epidural and 70 control patients. Nine allocated to epidural treatment were excluded after randomisation because of consent withdrawal, ineligibility or loss to follow-up; four controls were excluded for consent withdrawal or ineligibility. Although termed intention-to-treat by the investigators, this was a modified intention-to-treat population.¹

Mean age was 58 versus 57 years, and mean APACHE II score 18.7 versus 17.7. Gallstones and alcohol were common causes. Median symptom duration before randomisation was two days in each group; over half had symptoms for at least 48 hours. Invasive ventilation was already required in 13/65 and 14/70 patients. Vasopressor or inotrope use was appreciably higher in the epidural group, at 25/65 versus 13/70.¹

Among those with recorded timing, median catheter placement occurred 119 minutes after randomisation. Median epidural duration was six days. The allocation flow diagram records three failed insertions among 74 patients assigned to intervention, while the per-protocol population excluded additional treatment deviations and one control crossover. Most early recorded sensory-block assessments were unevaluable, limiting confirmation of functional blockade. The regimen reduced systemic opioid requirements, without an evident improvement in recorded pain scores.^{1,4}

Median ventilator-free days were 30 (interquartile range 15–30) with epidural treatment and 30 (18–30) with control. The estimated difference in medians was zero days, with a 95% confidence interval from –3.33 to 3.33 days ($P=0.59$). The maximum value of 30 days occurred in 37/65 and 42/70 patients. Distribution-focused adjusted analyses also found no clear group difference.^{1,4}

By day 30, 6/65 epidural patients and 10/70 controls had died: relative risk 0.65 (95% confidence interval 0.25–1.68; $P=0.36$). Invasive ventilation was required in 27/65 versus 29/70 patients. Among these ventilated patients, median ventilation duration was 14 days (5–28) versus six days (2–13), with $P=0.02$ for the reported comparison. This secondary comparison concerns a subset defined by ventilation during follow-up.¹

Acute respiratory distress syndrome occurred in 15/65 versus 10/70 patients, and renal replacement therapy in 12 patients in each group. There was no clear difference in new organ failure, necrosectomy or infected peripancreatic necrosis. Feeding tolerance, enteral calorie delivery and the measured biomarkers did not show a convincing benefit. Procedural events included vascular and dural puncture, a vasovagal event and mild transient hypotension. No epidural haematoma or serious epidural infection was identified. Exploratory subgroup findings did not establish a reliable treatment-responsive population.^{1,4}

Critique

The central difficulty is the mismatch between the expected primary-outcome distribution and the population actually recruited. The calculation anticipated a mean of 13 ventilator-free days and sought a seven-day improvement. In practice, more than half of each group reached the ceiling of 30 days. These patients had no scope for improvement on this endpoint, even if pain or some other aspect of recovery improved. Conversely, substantial deterioration would be needed to move many of them away from that ceiling. The nominal recruitment target therefore does not demonstrate that the trial retained its planned ability to detect the effect of interest.^{1,4}

The endpoint itself remains clinically meaningful: survival and freedom from invasive ventilation matter. Its limitations arise from the particular population and distribution. A zero value combines death with prolonged ventilation, while a maximum value incorporates the study's exemption for brief surgical ventilation. The 30-day maximum should not simply be translated into “never intubated”. The reported confidence interval around a difference in medians describes that chosen summary; it does not rule out every clinically important effect on mortality or on the more severely ill minority. EPIPAN is a failed superiority comparison, not an equivalence trial.^{1,4}

The ventilation-duration signal deserves attention without being converted into a definitive harm estimate. Only patients ventilated during follow-up enter that comparison. Treatment could influence who is ventilated, who survives and how long a survivor remains at risk of prolonged ventilation. Conditioning on this subsequent event weakens the protection afforded by randomisation. Fewer deaths in the epidural group could also affect observed duration, although the available data do not establish that explanation. The result is a legitimate safety concern requiring clarification, rather than proof that adding an epidural prolongs ventilation by eight days for the whole eligible population.¹

Precision and multiplicity reinforce that caution. Ventilation duration was one of many secondary outcomes. The supplementary and main reports examine numerous clinical, laboratory and subgroup comparisons without a confirmatory multiplicity framework for these findings. Table 2 gives a confidence interval crossing zero for the absolute difference in ventilation-duration medians, alongside a statistically significant comparison on another scale. These are not interchangeable analyses. The mortality interval permits substantial benefit and appreciable harm; 16 deaths cannot settle a survival claim. Selectively privileging either the favourable mortality point estimate or the unfavourable ventilation P value would overstate the evidence.¹

Randomisation and concealed allocation are important safeguards, but they did not produce closely matched circulatory support at baseline. Vasopressor or inotrope use was 38% versus 19%, a material imbalance for an intervention involving sympathetic blockade and for outcomes related to organ failure. This is compatible with chance in a small trial, not evidence of defective allocation. Nevertheless, it complicates interpretation of exploratory harm signals. Adjustment can help address measured differences, but cannot manufacture the information that a larger, adequately balanced sample would provide.¹

Post-randomisation exclusions also matter. Thirteen of 148 allocated patients were absent from the principal analysis, with more excluded from the epidural group. Some exclusions concerned consent and eligibility, but the loss of an allocated patient to follow-up is especially relevant when death determines the primary outcome. The retained comparison no longer includes everyone as randomised. Furthermore, the supplementary table labelled per-protocol repeats the primary-analysis denominators in its overall row despite the flow diagram identifying 61 and 69 per-protocol patients. The published table therefore does not permit independent confirmation of the claimed supporting per-protocol result.^{1,4}

Treatment delivery introduces a different uncertainty. Catheters were generally placed promptly after allocation and maintained for several days, so the strategy was more than a token exposure. However, many sensory-block assessments were unevaluable. Epidural local anaesthetic administration, demonstrated segmental blockade and improved pancreatic perfusion are separate steps. EPIPAN documented the first more reliably than the latter two. Its systemic opioid reduction supports a pharmacological effect, but the epidural solution itself contained sufentanil. It should not be described as an opioid-free strategy or as proof of reduced total opioid exposure across all routes.⁴

Timing and selection limit the biological claim further. More than half the cohort had symptoms for at least 48 hours, and eligibility did not require established severe organ failure. If the intended benefit depended on very early restoration of pancreatic microcirculation, this trial does not isolate that window. Equally, patients with refractory shock or major coagulation abnormalities were excluded. The observed results cannot justify extending the procedure to precisely those unstable patients in whom its haemodynamic and neuraxial risks may be greatest. These are limits of the tested population, not evidence that a different timing or severity group would benefit.¹

Open treatment assignment was unavoidable without a sham procedure, but its consequences differ by outcome. Death is comparatively resistant to observer interpretation, while pain treatment, sedation, intubation and weaning can respond to clinicians' knowledge of allocation. Blinded final assessment cannot remove differences introduced during care. The pragmatic design improves relevance to participating centres, yet variable co-interventions make it difficult to separate direct organ protection from altered analgesic or ventilatory management. Similar recorded pain scores do not establish formal analgesic non-inferiority, particularly with incomplete assessments.^{1,4}

Finally, reassuring absence of catastrophic neuraxial events in this small, selected sample is not a reliable estimate of rare-event safety. The technical requirements, anticoagulation constraints and difficulty assessing blockade remain relevant to implementation. Taken together, EPIPAN provides useful randomised evidence against routine adoption for organ protection, while leaving a narrower analgesic role unresolved. Its strongest practical message is that reduced systemic opioid use has not been shown here to translate into better major clinical outcomes.^{1,4}

Body of Evidence

The evidence developed along three distinct lines: experimental organ protection, clinical physiological studies and observational outcome associations. These should not

be treated as mutually confirming estimates of the same treatment effect. They ask different questions in different populations. EPIPAN is valuable because it brings the proposed organ-protective strategy into a multicentre randomised clinical comparison, where a plausible mechanism must produce a measurable patient benefit.

In 2013, Bachmann and colleagues randomised 34 pigs with experimentally induced severe pancreatitis to thoracic epidural bupivacaine or no epidural. Seven-day survival was 82% versus 29%, accompanied by improved pancreatic microcirculation, tissue oxygenation and histological injury. This provided a coherent mechanistic signal. However, the timing of induced disease and treatment, verified catheter position and controlled experimental conditions differ substantially from patients presenting after variable symptom durations. The experiment supported a human trial; its survival effect was not a credible prediction of the size of benefit that clinical practice should expect.²

Sadowski and colleagues subsequently studied epidural treatment in a small randomised clinical trial published in 2015. Thirty-eight patients were allocated and 35 analysed, 13 with epidural analgesia and 22 with intravenous patient-controlled analgesia. Improvement in pancreatic perfusion was more frequent with epidural treatment, but the numerical comparison involved measurements from pancreatic regions: 13/30 versus 2/27 measurements improved, rather than those numbers of patients. Intubation occurred in one versus six patients, without a statistically significant difference, and there were no deaths. The study supported feasibility and a physiological effect, while providing little precision for clinical outcomes. Its unit of perfusion measurement is particularly important when translating the apparent strength of the result into patient-level confidence.³

The 2018 multicentre propensity analysis by Jabaudon and colleagues included 1,003 patients from 17 French and Belgian intensive care units, of whom only 46 received an epidural. Thirty-day mortality was 4% with epidural treatment and 22% without it; matched analyses also favoured epidural treatment. The published abstract describes adjustment and propensity matching, but these techniques cannot remove differences in unmeasured suitability for neuraxial treatment. Selection for an invasive procedure may identify patients with a more favourable risk profile in ways not captured by recorded severity scores. This large association justified testing the hypothesis; it did not establish causality. EPIPAN's mortality results do not reproduce a sufficiently precise effect to validate that observational promise.⁵

Tyagi and colleagues' 2019 preliminary randomised study allocated 32 patients with predicted severe pancreatitis to receive or not receive a segmental thoracic epidural block. The published abstract found no statistically significant difference in the

aggregate Sequential Organ Failure Assessment score, mechanical ventilation requirement or hospital stay. Procalcitonin fell more with epidural treatment, but an isolated biomarker change did not establish improved organ function or survival. This small study therefore added another reason to distinguish physiological or inflammatory signals from patient benefit. It did not resolve the clinical uncertainty that EPIPAN was designed to address.⁶

The sequence explains why EPIPAN's neutral primary result changes the balance of evidence even though it is not definitive. The earlier case for disease modification rested on a striking animal effect, a small perfusion study and a susceptible observational comparison. Randomisation in a broader clinical setting did not reveal the anticipated improvement in freedom from ventilation. Nor did the measured clinical or biological secondary outcomes supply a consistent alternative signal of organ protection. The reduction in systemic opioid requirements remains a narrower finding with a different clinical purpose.^{1,4}

By the end of 2023, the unresolved question was therefore not whether an epidural had a plausible mechanism, but whether a selected analgesic indication or a better-defined treatment window could deliver a favourable benefit–risk balance. Establishing that would require reliable confirmation of blockade, a population whose relevant outcomes can improve, and adequate ascertainment of haemodynamic and neuraxial harms. Existing subgroup findings do not identify such a population. Routine epidural placement solely to modify the course of acute pancreatitis is not supported by this evidence.

Should thoracic epidural analgesia be added routinely to improve outcomes in intensive care patients with acute pancreatitis?

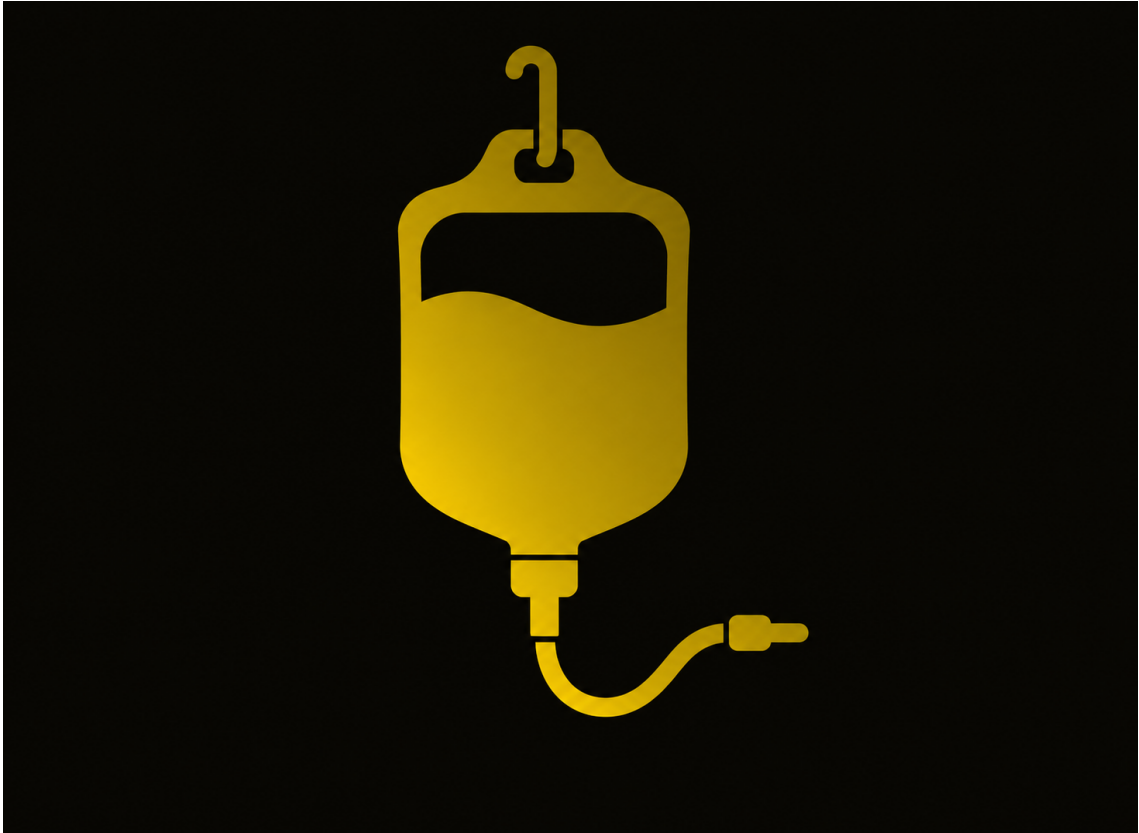
Not routinely. EPIPAN did not improve ventilator-free days, and neither an organ-protective benefit nor the safety of wider use has been established.

References

1. Jabaudon M, Genevrier A, Jaber S, Windisch O, Bulyez S, Laterre PF, et al.; EPIPAN study group. Thoracic epidural analgesia in intensive care unit patients with acute pancreatitis: the EPIPAN multicenter randomized controlled trial. *Crit Care* 2023;27:213.
2. Bachmann KA, Trepte CJC, Tomkötter L, Hinsch A, Stork J, Bergmann W, et al. Effects of thoracic epidural anesthesia on survival and microcirculation in severe acute pancreatitis: a randomized experimental trial. *Crit Care* 2013;17:R281.

3. Sadowski SM, Andres A, Morel P, Schiffer E, Frossard JL, Platon A, et al. Epidural anesthesia improves pancreatic perfusion and decreases the severity of acute pancreatitis. *World J Gastroenterol* 2015;21:12448–12456.
4. EPIPAN study group. Additional file 1 to: Thoracic epidural analgesia in intensive care unit patients with acute pancreatitis: the EPIPAN multicenter randomized controlled trial. *Crit Care* 2023;27:213.
5. Jabaudon M, Belhadj-Tahar N, Rimmelé T, Joannes-Boyau O, Bulyez S, Lefrant JY, et al.; Azurea Network. Thoracic Epidural Analgesia and Mortality in Acute Pancreatitis: A Multicenter Propensity Analysis. *Crit Care Med* 2018;46:e198–e205.
6. Tyagi A, Gupta YR, Das S, Rai G, Gupta A. Effect of Segmental Thoracic Epidural Block on Pancreatitis-induced Organ Dysfunction: A Preliminary Study. *Indian J Crit Care Med* 2019;23:89–94.

Possible Reporting Inconsistencies. Supplementary Table 19, labelled per protocol, retains overall groups of 70 controls and 65 epidural recipients, whereas the flow diagram gives 69 and 61; its strata also total 130 rather than 135. That table cannot be treated as an independently verified per-protocol confirmation. In addition, the prose confidence interval for the ventilation-duration coefficient, 0.56 (0.11–0.55), excludes its own estimate; the upper limit appears mistyped, with Table 2 giving 1.02.^{1,4}



Nutrition Trials

TGC-FAST

Gunst J, Debaveye Y, Güiza F, Dubois J, De Bruyn A, Dauwe D, et al.; TGC-Fast Collaborators. Tight Blood-Glucose Control without Early Parenteral Nutrition in the ICU. *N Engl J Med* 2023;389(13):1180–1190.

Introduction

Glucose control in intensive care illustrates the distance between correcting an abnormal measurement and improving a patient's outcome. Hyperglycaemia accompanies critical illness and is associated with adverse outcomes, but this does not establish that lowering glucose into the normal fasting range is beneficial. The intervention brings its own risks, particularly hypoglycaemia, and depends on how insulin is administered, glucose is measured and nutrition is delivered.¹

The early enthusiasm for intensive insulin treatment followed a single-centre surgical trial in which tight control reduced intensive care mortality. A later large international trial, NICE-SUGAR, found increased 90-day mortality with intensive control. The clinical question consequently became more specific: was normalisation intrinsically unhelpful, or had variation in treatment delivery and concomitant nutrition changed its benefit–risk balance?^{2,3}

Two developments made this question worth revisiting. Computer-assisted insulin dosing offered more consistent glucose control, while evidence supporting delayed supplemental parenteral nutrition changed the metabolic context in which insulin was being used. Avoiding early parenteral calories might reduce hyperglycaemia and therefore the opportunity for benefit from aggressive correction. Equally, lower nutrient delivery could make safe insulin titration more demanding. These possibilities cannot be resolved by comparing glucose concentrations across unrelated studies.^{1,4,5}

TGC-FAST tested a deliberately integrated strategy: tight glucose control using the LOGIC-Insulin algorithm, compared with a liberal insulin policy, while withholding early supplemental parenteral nutrition in both groups. The investigators anticipated a smaller benefit than in earlier work and recruited a correspondingly large sample. The trial asked whether normal fasting glucose could be achieved with little severe hypoglycaemia and whether doing so would shorten the period during which intensive care was required. It was a renewed test of clinical efficacy under specified nutritional and monitoring conditions, not simply a demonstration that software could lower glucose.¹

Synopsis

TGC-FAST was an investigator-initiated, multicentre, parallel-group randomised trial in 11 intensive care units across three Belgian hospitals: two university hospitals and one district hospital. Adults were recruited from September 2018 to August 2022. Recruitment paused during the first COVID-19 wave and subsequently excluded patients in closed COVID-19 units. Funding came from public research bodies, university and hospital sources, and a private donation. The investigators designed and analysed the trial.¹

Patients were enrolled on intensive care admission. Important exclusions included absence of the vascular access required for treatment and monitoring, an expectation of eating, treatment limitations or a moribund condition, previous participation, anticipated early parenteral nutrition or high-dose glucose, and prolonged admission to a non-participating intensive care unit. Diabetic ketoacidosis and hyperosmolar hyperglycaemic states were excluded. Written informed consent was obtained from the patient or a legal representative.¹

Central computer randomisation allocated patients equally, in blocks of ten, stratified by centre and admission diagnostic category. Bedside staff did not know the block size, but treatment assignment was open. Tight control targeted glucose at 80–110 mg/dL (4.4–6.1 mmol/L). LOGIC-Insulin advised the insulin dose, glucose rescue dose and timing of the next measurement, normally every one to four hours and more frequently after hypoglycaemia.¹

With liberal control, insulin began after two consecutive glucose measurements above 215 mg/dL (11.9 mmol/L), or one such measurement in patients with type 1 diabetes. Once required, insulin was titrated by bedside staff to 180–215 mg/dL (10.0–11.9 mmol/L). Electronic alerts indicated when insulin should be started or discontinued. Glucose was measured at least four times daily, and hypoglycaemia treatment followed local clinical judgement. Thus, the comparator was an active treatment policy permitting spontaneous lower glucose values, rather than an instruction to raise glucose into a high range.¹

Both groups used arterial samples analysed on blood-gas instruments and continuous intravenous insulin through a central venous catheter. Capillary measurements were permitted after arterial-line removal. Assigned treatment stopped when patients began eating, lost central venous access or left intensive care, whichever came first; it resumed after readmission within 48 hours. Enteral feeding began as soon as possible. Supplemental parenteral nutrition was deferred until after one week, with parenteral micronutrients according to local practice when enteral intake was inadequate.¹

The primary outcome was time to discharge alive from intensive care, using readiness for discharge or actual discharge, whichever occurred first. Readiness meant that patients were no longer at risk of, or in need of, vital-organ support. Death was treated as a competing event by retaining nonsurvivors without a discharge event until a censoring time beyond the last surviving patient's observation. A higher discharge hazard favoured faster recovery. The principal safety outcome was death within 90 days.¹

Severe hypoglycaemia meant glucose below 40 mg/dL (2.2 mmol/L); resistance to intravenous glucose rescue constituted a serious adverse event. Eight prespecified secondary outcome categories covered kidney injury, liver dysfunction, new infections, respiratory and haemodynamic support, live hospital discharge, intensive care mortality and hospital mortality. Severe acute kidney injury used stage 3 creatinine criteria; new kidney-replacement therapy was also assessed. Liver outcomes were biochemical thresholds, rather than a single clinical liver-failure endpoint.¹

The planned sample of 9,230 provided 80% power at an alpha level of 0.05 to detect a one-day difference in intensive care requirement and to exclude a 1.5-percentage-point mortality increase. An independent board reviewed duration assumptions after 25% recruitment and mortality after 50%; both reviews allowed continuation. The primary time-to-event comparison used a Cox model, with the proportional-hazards assumption assessed. All randomised patients formed the intention-to-treat population. Secondary and subgroup confidence intervals were not adjusted for multiple comparisons.¹

Of 24,195 assessed patients, 9,230 were randomised: 4,608 to tight and 4,622 to liberal control. All remained in the trial analysis, although day-90 vital status was unavailable for one patient in each group. Median age was 67 years and median APACHE II score 21 in both groups. About one fifth had diabetes, 28% had sepsis and 45% were admitted after cardiac surgery or its complications. Baseline characteristics were closely balanced.¹

Insulin was administered to 98.8% of tight-control patients and 45.9% of liberal-control patients. Median patient-level average daily insulin dose was 24.8 versus zero units. Median patient-level average morning glucose was 107 versus 140 mg/dL, while corresponding all-day values were 115 versus 145 mg/dL. Morning measurements were unavailable for 91 patients. Recorded parenteral calorie delivery in the first week was low and similar, with a median of 101 kcal/day in each group; it was not literally zero.^{1,6}

Time to live intensive care discharge was similar: hazard ratio 1.00 (95% confidence interval 0.96–1.04; $P=0.94$). Median intensive care duration was three days in both groups. By day 90, 486/4,607 tight-control patients and 468/4,621 liberal-control patients had died, 10.5% versus 10.1%: relative risk 1.04 (0.92–1.17; $P=0.51$). Severe

hypoglycaemia occurred in 47/4,608 versus 31/4,622 patients, 1.0% versus 0.7%: relative risk 1.52 (0.97–2.39). No episode was resistant to treatment.¹

Severe acute kidney injury occurred in 326/4,507 versus 387/4,500 patients, relative risk 0.84 (0.73–0.97), and new kidney-replacement therapy in 212/4,507 versus 259/4,500, relative risk 0.82 (0.68–0.98). These analyses excluded 223 patients already receiving kidney replacement or chronic dialysis. Some cholestatic enzyme thresholds were less frequently reached with tight control, without a clear effect on bilirubin or cytolytic markers. New infection, organ-support duration, hospital discharge and other mortality outcomes showed no clear differences.^{1,6}

Six prespecified subgroup comparisons included diabetes, diagnostic groups and predicted prolonged intensive care stay. Only neurological or neurosurgical admission suggested possible mortality heterogeneity: 47/525 versus 66/528 deaths, with an odds ratio of 0.69 (0.46–1.02). The primary discharge outcome did not show a corresponding advantage. The predicted-stay subgroup used an admission-variable model developed in liberal-control patients and evaluated in tight-control patients.^{1,6}

Critique

TGC-FAST provides a strong answer to its main clinical question because the intervention was successfully delivered and the primary estimate is precise. Glucose was appreciably lower with tight control, insulin exposure was markedly greater, and the discharge hazard ratio was centred on one with a narrow confidence interval. Failure to separate treatment groups is therefore an implausible explanation for the absence of benefit. Under these conditions, normalising glucose did not produce a detectable overall acceleration in intensive care recovery.¹

The achieved comparator is crucial. The liberal policy allowed insulin to be withheld until glucose exceeded 215 mg/dL, but its median morning glucose was only 140 mg/dL. TGC-FAST therefore does not compare sustained normal glucose with sustained glucose around 200 mg/dL in most participants. Nor does it show that persistently severe hyperglycaemia can be ignored. Its practical comparison is the additional effect of algorithm-driven normalisation in a population whose glucose was often moderately elevated without insulin. The treatment thresholds and the observed concentrations answer different questions and should not be interchanged.¹

Nutrition was controlled sufficiently to frame that question, but was not itself randomised. Both groups deferred supplemental parenteral nutrition and received similar low first-week parenteral calorie exposure. This protects the insulin comparison

from a major between-group nutritional imbalance. It cannot establish that delaying nutrition improved either group compared with early feeding, or prove that nutrition caused the disagreement between historical insulin trials. The small recorded calorie intake also prevents a literal interpretation of the trial as involving no intravenous calories whatsoever.^{1,6}

The safety result is reassuring relative to the frequency of severe hypoglycaemia in earlier intensive-control studies, but it is not evidence that hypoglycaemia was abolished. Forty-seven tight-control patients experienced a glucose value below 40 mg/dL, and the confidence interval around the relative risk permits a substantial increase. The absence of treatment-resistant episodes answers a narrower question. Different sampling frequencies may also affect event detection. These findings belong to a package involving software, trained staff, frequent measurements and predominantly arterial analysis; reproducing only its glucose target would not reproduce its safety conditions.¹

Mortality requires a similarly exact interpretation. The study was sized with the aim of excluding a 1.5-percentage-point increase, yet the publication presents a nonsignificant superiority comparison and a relative-risk interval. A simple independent calculation from the reported counts gives an absolute difference of 0.42 percentage points, with an approximate unadjusted 95% confidence interval from -0.82 to 1.66 percentage points. This calculation uses the usual normal approximation and is not a reported trial estimate. It illustrates why absence of statistical significance is insufficient to declare that the envisaged safety margin has been met.¹

The primary endpoint addresses a relevant burden of illness while incorporating an important safeguard: death cannot appear beneficial merely by shortening observed stay. Nevertheless, readiness for discharge contains clinical judgement. Staff knew allocation, and the study crossed a pandemic during which bed availability and discharge practice changed. Using readiness rather than physical transfer mitigates some logistical delay, but cannot guarantee uniform decisions. Randomisation within centres helps distribute these pressures; the close agreement of mortality, hospital discharge and organ-support outcomes makes a large hidden global benefit less plausible.¹

The kidney results are the most credible reason to avoid interpreting the trial as showing no biological effect of any kind. Severe creatinine-defined injury and new kidney-replacement therapy both favoured tight control. However, these are related outcomes, not independent replications, and they sit among multiple unadjusted secondary comparisons. Kidney-replacement initiation also involves clinical decisions in an open trial. The signal supports further investigation of renal effects, but does not

overturn the primary result or establish that treating every eligible patient intensively has a favourable overall balance.¹

The liver findings warrant still greater restraint. Changes in the frequency of enzyme thresholds do not necessarily represent prevention of clinically important liver failure. Different numbers of patients had the relevant measurements, and bilirubin and cytolytic markers did not provide consistent confirmation across the entire outcome category. A reduction in selected laboratory abnormalities can be real without improving survival, recovery or symptoms. It should therefore remain a secondary physiological observation rather than a substitute primary success.^{1,6}

Subgroup interpretation is another potential source of overreach. The neurological mortality finding arose among several prespecified comparisons, its own odds-ratio interval included one, and there was no parallel improvement in the primary endpoint. Prespecification helps credibility but does not remove multiplicity. The predicted-long-stay analysis is conceptually preferable to selecting patients by their actual subsequent stay, which could be altered by treatment. However, the prediction model was developed within this trial, and fewer than half of those classified as likely long stayers actually stayed beyond seven days. It does not provide a validated rule for selecting patients for intensive insulin.^{1,6}

Generalisability depends more on the composition and delivery of care than on the large headline sample. Almost half the cohort followed cardiac surgery, most belonged to surgical or trauma groups, and median intensive care duration was only three days. The evidence is less direct for populations dominated by prolonged medical critical illness, closed COVID-19 units, hyperglycaemic emergencies, or settings without comparable monitoring infrastructure. Excluding patients expected to eat and those without vascular access also defines the treatment context. These qualifications do not undermine internal validity; they specify where applying the result is most defensible.¹

The overall judgement is therefore proportionate but clinically consequential. TGC-FAST substantially weakens the case for routine normalisation of glucose in similarly managed adults receiving no early supplemental parenteral nutrition. It demonstrates that improving control and limiting severe hypoglycaemia can coexist with a neutral clinical outcome. The renal and neurological signals preserve focused research questions, while providing insufficient grounds for a general change to tight control.¹

Body of Evidence

The 2001 Leuven surgical trial established the original case for intensive insulin treatment. It randomised 1,548 patients to glucose of 80–110 mg/dL or conventional insulin treatment initiated above 215 mg/dL. Intensive care mortality was 4.6% versus 8.0%, with a sequentially adjusted P value below 0.04. This was a major clinical signal, not merely improved glycaemia. Its single-centre surgical setting and the early nutritional practices described in the later TGC-FAST report are central to interpreting its applicability. A subsequent neutral result under different nutritional conditions cannot identify which individual feature explains the difference.^{1,2}

The 2006 Leuven medical trial extended the question to 1,200 patients expected to require at least three days of intensive care. Its principal intention-to-treat comparison did not significantly reduce in-hospital mortality, despite improvements in some morbidity measures. This already complicated a universal claim for tight control. The index report's broader historical account of beneficial early trials should therefore be read with the distinction between mortality and morbidity intact. Not every early trial showed an overall survival benefit.⁷

NICE-SUGAR provided a larger, international test in 2009. Among 6,104 randomised adults expected to need at least three days of intensive care, intensive control targeted 81–108 mg/dL and conventional control no more than 180 mg/dL. Ninety-day mortality was 27.5% versus 24.9%, with an odds ratio of 1.14 (95% confidence interval 1.02–1.28). Severe hypoglycaemia was substantially more frequent with intensive treatment. This established harm from that intensive strategy in its tested population. TGC-FAST's lower absolute hypoglycaemia incidence helps distinguish the delivery packages, but comparing two trials cannot establish that hypoglycaemia was the sole cause of NICE-SUGAR's mortality difference.³

EPaNIC then changed the nutritional context. In 2011, 4,640 adults were randomised to supplemental parenteral nutrition within 48 hours or no earlier than day eight, with early enteral feeding and glucose normalisation in both groups. Delayed supplementation modestly increased the likelihood of earlier live intensive care discharge, hazard ratio 1.06 (1.00–1.13), and reduced infections without a clear mortality difference. This was direct evidence about nutrition timing. It supplied a rationale for TGC-FAST's common feeding strategy, while leaving the incremental benefit of glucose normalisation in that setting unanswered.⁵

LOGIC-2 addressed treatment delivery rather than the optimum target. In 2017, 1,550 patients were randomised to software-guided or expert nurse-directed insulin, using each centre's chosen glucose range. Software improved the primary glycaemic penalty

score and time within target; severe hypoglycaemia occurred in 0.9% versus 1.2%. Clinical outcomes did not clearly differ. This demonstrated better control of a measurement, not improved survival from a lower target. TGC-FAST supplied the subsequent large clinical test of combining this technology with normal fasting glucose targets.⁴

SHINE provides relevant, though indirect, context for the neurological subgroup. In 2019, 1,151 patients with acute ischaemic stroke and hyperglycaemia were randomised to intensive intravenous insulin targeting 80–130 mg/dL or standard subcutaneous treatment targeting 80–179 mg/dL for up to 72 hours. The trial stopped for futility: favourable 90-day functional outcome occurred in 20.5% versus 21.6%, adjusted relative risk 0.97 (0.87–1.08). Severe hypoglycaemia occurred only with intensive treatment. Its population and endpoint differ from TGC-FAST's mixed neurological category, but it argues against converting an exploratory mortality signal into a broad neurological indication.⁸

Together, these studies separate three propositions that were initially easy to conflate: hyperglycaemia predicts risk, software can improve glycaemic control, and normalising glucose improves clinical outcomes. The first two do not establish the third. By the end of 2023, TGC-FAST added a large, well-delivered neutral comparison in a setting that deferred early parenteral nutrition. It narrows the explanation that earlier neutral or harmful results were simply failures of insulin administration, while leaving the exact optimum between liberal treatment thresholds and normal fasting targets unresolved. Its results support restraint in routine normalisation, with confirmatory work needed before acting on organ-specific or subgroup findings.¹

Should normal fasting glucose be targeted routinely in critically ill adults receiving no early supplemental parenteral nutrition?

No. Algorithm-guided tight control did not improve intensive care recovery or survival, and the secondary signals do not justify routine adoption.

References

1. Gunst J, Debaveye Y, Güiza F, Dubois J, De Bruyn A, Dauwe D, et al.; TGC-Fast Collaborators. Tight Blood-Glucose Control without Early Parenteral Nutrition in the ICU. *N Engl J Med* 2023;389(13):1180–1190.
2. Van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schetz M, et al. Intensive insulin therapy in critically ill patients. *N Engl J Med* 2001;345:1359–1367.

3. Finfer S, Chittock DR, Su SY, Blair D, Foster D, Dhingra V, et al.; NICE-SUGAR Study Investigators. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med* 2009;360:1283–1297.
4. Dubois J, Van Herpe T, van Hooijdonk RT, Wouters R, Coart D, Wouters P, et al. Software-guided versus nurse-directed blood glucose control in critically ill patients: the LOGIC-2 multicenter randomized controlled clinical trial. *Crit Care* 2017;21:212.
5. Casaer MP, Mesotten D, Hermans G, Wouters PJ, Schetz M, Meyfroidt G, et al. Early versus late parenteral nutrition in critically ill adults. *N Engl J Med* 2011;365:506–517.
6. TGC-Fast Collaborators. Supplementary Appendix to: Tight Blood-Glucose Control without Early Parenteral Nutrition in the ICU. *N Engl J Med* 2023;389:1180–1190.
7. Van den Berghe G, Wilmer A, Hermans G, Meersseman W, Wouters PJ, Milants I, et al. Intensive insulin therapy in the medical ICU. *N Engl J Med* 2006;354:449–461.
8. Johnston KC, Bruno A, Pauls Q, Hall CE, Barrett KM, Barsan W, et al.; Neurological Emergencies Treatment Trials Network and SHINE Trial Investigators. Intensive vs Standard Treatment of Hyperglycemia and Functional Outcome in Patients With Acute Ischemic Stroke: The SHINE Randomized Clinical Trial. *JAMA* 2019;322:326–335.

NUTRIREA-3

Reignier J, Planteveve G, Mira JP, Argaud L, Asfar P, Aissaoui N, et al.; NUTRIREA-3 Trial Investigators and Clinical Research in Intensive Care and Sepsis (CRICS-TRIGGERSEP) Group. Low versus standard calorie and protein feeding in ventilated adults with shock: a randomised, controlled, multicentre, open-label, parallel-group trial (NUTRIREA-3). *Lancet Respir Med* 2023;11(7):602–612.

Introduction

Early nutritional support in critical illness has competing objectives. Clinicians want to limit cumulative energy deficits and loss of lean tissue, while avoiding a metabolic load that an acutely ill patient cannot use effectively. Shock adds uncertainty: gastrointestinal perfusion may be impaired, feeding intolerance is common, and intravenous nutrition introduces a different set of practical and metabolic considerations. A prescription intended to replace calculated requirements therefore needs evidence that its timing and intensity improve recovery.¹

Calories and protein also raise different questions. Reducing energy delivery may limit overfeeding, but reducing amino acid provision could theoretically aggravate muscle loss. Conversely, providing more substrate does not establish that it is incorporated into functional tissue during the acute response to illness. Nutritional interventions can change fluid administration, insulin requirements and the frequency of gastrointestinal complications alongside nutrient intake. These effects matter when interpreting a clinical benefit attributed to feeding intensity.¹

Earlier randomised studies had challenged immediate pursuit of full nutritional targets. EDEN compared trophic with full enteral feeding in acute lung injury, while PermiT reduced nonprotein calories with similar protein provision between groups. Neither established a survival advantage from the higher calorie strategy. These comparisons did not directly answer whether restricting both calories and protein during the earliest phase of ventilated shock could improve recovery.^{2,3}

NUTRIREA-3 addressed that specific question. It compared a markedly restricted regimen with standard calorie and protein targets during the first week of intensive care, allowing enteral or parenteral administration according to clinical circumstances. Its two primary outcomes deliberately captured different dimensions of benefit: survival to 90 days and readiness to leave intensive care.¹

Synopsis

NUTRIREA-3 was a pragmatic, open-label, parallel-group superiority trial in 61 French intensive care units, including overseas territories. Recruitment ran from 5 July 2018 to 8 December 2020. Nantes University Hospital sponsored the study, which received French Ministry of Health funding; the funder had no role in its conduct, analysis or reporting. The trial was registered as NCT03573739.¹

Eligible adults were invasively ventilated, expected to remain so for at least 48 hours, and receiving adrenaline, dobutamine or noradrenaline for shock. The timing criteria selected patients within 24 hours of intubation in intensive care, or admitted within 24 hours after intubation elsewhere. Nutrition was intended to start within the corresponding early window. Important exclusions included established long-term home enteral or parenteral nutrition or specific nutritional requirements, impending death or treatment limitation, pregnancy or recent childbirth, lactation, and specified legal restrictions. Emergency enrolment was possible when the patient lacked capacity and relatives were unavailable; subsequent refusal required exclusion.¹

Central computer-generated allocation assigned patients equally to the two strategies, using variable blocks and stratification by centre. The secure web system concealed the allocation sequence, but bedside clinicians and nurses knew treatment assignment. An independent masked committee adjudicated intensive care infections; other clinical assessments were not thereby blinded.¹

For the first seven days, the low group targeted 6 kcal/kg/day and 0.2–0.4 g/kg/day of protein. The standard group targeted 25 kcal/kg/day and 1.0–1.3 g/kg/day. Weight was adjusted for obesity and underweight: a weight corresponding to body mass index 30 kg/m² was used above that threshold, while the midpoint of actual and ideal weight was used below 18.5 kg/m². Commercial formulations supplied the assigned nutrients without separate protein supplements. Feeding began at the full assigned rate, delivered continuously over 24 hours, rather than through a common gradual escalation schedule.¹

The attending physician selected enteral or parenteral administration daily. The assigned regimen continued until both extubation and withdrawal of vasoactive treatment, death, or completion of day seven. Reintubation during that period prompted resumption during the acute phase. From day eight, both groups targeted 30 kcal/kg/day and 1.2–2.0 g/kg/day of protein. Fluids, electrolytes, micronutrients and glucose management remained responsive to clinical needs and local practice. This was a temporary early restriction strategy, not prolonged low-protein feeding.¹

The coprimary outcomes were all-cause mortality at 90 days and time to readiness for intensive care discharge. Readiness required freedom from a need, or impending need, for invasive ventilation and vasoactive drugs, together with absence of neurological, metabolic or haematological problems requiring close intensive care management. Bedside physicians assessed this daily, independently of bed availability. A Fine–Gray model treated death in intensive care as a competing event; a subdistribution hazard ratio above one favoured earlier readiness. Secondary outcomes included liberation from support, infections, gastrointestinal events, biochemical disturbances and measures of strength and mobilisation.^{1,4}

The planned sample of 3,044 provided 80% power to detect a five-percentage-point mortality reduction from an assumed 43% to 38%. It also provided 94% power for a 1.5-day difference in discharge readiness, from 14.5 to 13 days. Two interim analyses, after 1,000 and 2,000 patients, used a significance threshold of 0.001; the final threshold was 0.049. Both reviews recommended continuation. There was no additional adjustment for the two primary outcomes or the numerous secondary comparisons. Three patients with missing 90-day vital status were assumed dead in the primary analysis.¹

Of 10,767 patients assessed, 3,044 were randomised: 1,525 to low and 1,519 to standard feeding. Eight subsequent consent refusals, four per group, left 1,521 and 1,515 for analysis. Besides 5,507 patients meeting exclusion criteria, 2,216 eligible patients were not enrolled, commonly because of consent difficulties, oversight or logistical constraints. Mean age was 66 years in each arm and approximately two-thirds were men. More than 80% had medical admissions, approximately 58% had sepsis, median Sequential Organ Failure Assessment score was 10, and roughly 90% had no pre-existing malnutrition by the study classification.^{1,4}

The interventions began early and produced substantial separation. Feeding started a median 16.7 versus 17.0 hours after intubation. Median daily total energy delivery, including propofol and intravenous glucose, was 7.4 versus 22.0 kcal/kg, and protein delivery was 0.2 versus 0.9 g/kg. Enteral nutrition alone was used in 61.6% versus 56.0% of patients with available route data; parenteral nutrition alone in 23.7% versus 23.1%. Median cumulative fluid intake over days zero to seven was 11.1 versus 17.2 litres. This total included nutritional and other fluids and was not a measure of net fluid balance.^{1,4}

At 90 days, 628/1,521 patients assigned low feeding and 648/1,515 assigned standard feeding had died, 41.3% versus 42.8%. The absolute difference was –1.5 percentage points (95% confidence interval –5.0 to 2.0; $P=0.41$). Excluding the three patients with missing vital status did not materially change this result. Reported median time to

readiness for discharge was eight days (interquartile range 5–14) versus nine days (5–17), with a subdistribution hazard ratio of 1.12 (95% confidence interval 1.02–1.22; $P=0.015$).¹

Liberation from invasive ventilation occurred earlier with low feeding: reported medians were five versus six days, hazard ratio 1.12 (95% confidence interval 1.03–1.22; $P=0.007$). Vasoactive withdrawal had a median of three days in both groups and did not meet conventional statistical significance. Among survivors, median intensive care stay was nine versus ten days and hospital stay 21 versus 22 days. No significant difference in 28-day mortality was established.¹

The reported cumulative incidences of vomiting and diarrhoea were lower with restricted feeding, 20.2% versus 25.5% and 28.9% versus 33.3%, respectively. Bowel ischaemia was reported in 0.9% versus 1.8%, hazard ratio 0.50 (95% confidence interval 0.26–0.95; $P=0.030$). Intensive care infections were numerically less frequent, 15.3% versus 17.5%, but their interval included no effect; ventilator-associated pneumonia was similar. During days zero to seven, hypophosphataemia and insulin use were less frequent with low feeding. Hypoglycaemia occurred in 6.0% versus 4.9%, without a statistically significant difference.¹

At intensive care discharge, median Medical Research Council strength score was 49 versus 52 ($P=0.20$). The report did not establish improved strength or mobilisation with either regimen. Laboratory-defined liver dysfunction was less frequent with low feeding, but this endpoint encompassed biochemical abnormalities rather than a demonstrated reduction in clinical liver failure.^{1,4}

Critique

The most defensible interpretation is that a restricted early feeding package modestly accelerated recovery sufficiently to meet discharge-readiness criteria, without establishing a survival benefit. The two primary findings should remain distinct. Mortality was close to the anticipated control rate, and its confidence interval still permits a clinically important reduction as well as a smaller increase. A nonsignificant superiority test cannot establish equivalent survival or prove that severe early nutritional restriction is harmless.¹

The comparison also cannot isolate protein. Calories and protein changed together, and the low group received substantially less total fluid. Differences in feeding route emerged under clinician discretion. Randomisation supports a causal comparison of assignment to the two nutritional strategies, including their downstream consequences. It does not identify whether energy restriction, protein restriction, fluid exposure,

tolerance, or an interaction among these features caused earlier recovery. The fluid difference is particularly substantial, but its interpretation must remain disciplined: cumulative intake is not balance, and no randomised fluid comparison was performed.^{1,4}

Treatment separation was strong enough to make a neutral result informative. The standard arm achieved less than its nominal targets, especially for protein, yet received approximately three times the total calories and more than four times the protein delivered in the low arm. Failure to reach every prescribed target does not erase this contrast. Conversely, the trial does not compare low feeding with a carefully titrated, intermediate regimen. Starting the standard prescription immediately at full rate may matter as much as its eventual target when considering transfer to units that escalate feeding gradually.¹

Readiness for discharge is a sensible attempt to measure recovery without contamination from bed availability. Its specified criteria make it more interpretable than an unexplained judgement that a patient can leave intensive care. However, the outcome still depended on unblinded bedside decisions concerning ventilation, vasoactive support, neurological status and monitoring needs. Knowledge of feeding assignment could affect those decisions. Earlier ventilator liberation provides supporting coherence, but it too involves clinician action and shares part of the same recovery process. It is supportive evidence, not wholly independent confirmation.¹

The competing-risk analysis is preferable to simply discarding deaths or treating death as an ordinary discharge. Nevertheless, its hazard ratio describes the cumulative incidence process under that model. It does not mean every patient left intensive care 12% sooner, and it cannot be translated into a guaranteed one-day saving. The reported medians and survivor-only lengths of stay also answer different questions. In a population with more than 40% mortality by 90 days, maintaining these distinctions is essential to avoid making a modest recovery result appear to be an improvement in survival-adjusted long-term health.¹

The trial has important safeguards against systematic allocation bias. Central concealment, balanced groups, broad multicentre recruitment and a large sample support the reliability of the strategy comparison. Exclusion of four post-allocation consent withdrawals in each group means the analysed cohort was not literally every randomised patient, but the small, balanced loss is unlikely by itself to explain the recovery result. The three missing mortality outcomes were also too few to materially change the observed conclusion, as the sensitivity analysis demonstrated. These are concrete strengths rather than assurances derived from the trial's size alone.¹

Statistical interpretation requires proportional caution. Adjustment for interim examinations preserved most of the conventional significance threshold, but did not address testing two primary outcomes and many secondary endpoints. The discharge-readiness P value is stronger than a marginal result just below 0.05, so multiplicity alone is not a sufficient reason to dismiss it. Secondary signals nevertheless remain less secure, particularly uncommon events such as bowel ischaemia and biochemical endpoints with several related definitions. Their direction and plausibility can support the overall interpretation without turning each into a separately proven clinical benefit.¹

Gastrointestinal tolerance is a relevant patient and nursing outcome, but reduced vomiting or diarrhoea during a much smaller feeding load is not unexpected. These events may partly mediate recovery, reflect a simpler workload, or both. The bowel-ischaemia signal is more consequential, yet sparse events produce an imprecise estimate. Not every adverse event favoured restriction: colonic pseudo-obstruction was recorded in eight low-group versus two standard-group patients. Neither this imbalance nor the favourable ischaemia result alone establishes a comprehensive gastrointestinal safety profile.¹

The strength findings do not close the protein question. A nonsignificant discharge strength comparison is not a functional non-inferiority analysis, and the assessment denominator was not clearly reported. Patients who survive and can undertake strength testing are a selected subset of the original cohort. Early changes in weight and circulating nutritional markers are also difficult to interpret during inflammation and large fluid shifts. Without a robust longer-term functional assessment in the supplied report, the trial cannot establish that profound early protein restriction preserves muscle or later independence.^{1,4}

Applicability is strongest for promptly enrolled, invasively ventilated adults with shock resembling this predominantly medical French population. Most were not malnourished before admission. Patients requiring established specialist nutritional support, those presenting outside the early window and those with treatment limitations were excluded. The sizeable number of eligible but unenrolled patients further cautions against assuming universal representativeness. These features do not invalidate the result; they define the patients in whom its balance of benefit and uncertainty is most directly supported.¹

Finally, the intervention was explicitly time limited. After the acute treatment period, nutritional targets increased and became the same in both arms. Extending 6 kcal/kg/day and 0.2–0.4 g/kg/day protein indefinitely would be a different, untested policy. The

useful bedside lesson is to reconsider an immediate full early prescription in comparable shock, while reassessing nutritional needs as the acute phase resolves. It is not evidence for neglecting nutrition throughout a prolonged critical illness.¹

Body of Evidence

The earlier nutrition trials are best understood by the dimension each changed. Timing of supplemental parenteral nutrition, calorie dose, protein dose and route are related but distinct interventions. Their neutral survival results do not make those questions interchangeable. NUTRIREA-3 adds evidence about simultaneous early calorie and protein restriction in a particularly sick population; it should be placed beside the earlier comparisons without assuming that they all test the same treatment.

In EPaNIC, 4,640 adults were assigned early supplemental parenteral nutrition, within 48 hours, or withholding it until day eight when needed; both groups received early enteral nutrition. Later supplementation was associated with earlier live discharge and fewer intensive care infections, without a mortality advantage. Its relevance is the challenge to immediate completion of calculated nutritional requirements. It did not directly test the NUTRIREA-3 calorie and protein targets, and both EPaNIC groups also received intensive glucose management.⁵

EDEN randomised 1,000 ventilated patients with acute lung injury to trophic or full enteral feeding for six days. Intake averaged approximately 400 versus 1,300 kcal/day. Ventilator-free days to day 28 were similar, 14.9 versus 15.0, and 60-day mortality was 23.2% versus 22.2%; trophic feeding caused less gastrointestinal intolerance. This earlier pulmonary population supports the feasibility of an initially smaller enteral load. Its neutral primary outcome does not independently confirm faster recovery in ventilated shock.²

PermiT more clearly separated the calorie question from the protein question. Among 894 critically ill adults, permissive underfeeding targeted 40–60% of calculated energy requirements versus 70–100%, with similar protein intake. Achieved energy provision differed substantially, while protein averaged 57 versus 59 g/day. Ninety-day mortality was 27.2% versus 28.9%, relative risk 0.94 (95% confidence interval 0.76–1.16). These findings did not establish benefit from the higher calorie strategy, but they cannot be used to infer that NUTRIREA-3's very low protein target is independently beneficial.³

TARGET subsequently tested energy-dense versus routine enteral formula in 46 Australian and New Zealand intensive care units. Its modified intention-to-treat analysis included 3,957 ventilated adults. Energy intake averaged 1,863 versus 1,262 kcal/day;

90-day mortality was 26.8% versus 25.7%, relative risk 1.05 (95% confidence interval 0.94–1.16). More enteral calories again failed to improve survival. This result, published after NUTRIREA-3 recruitment began, belongs to the evidence available for interpreting the completed trial rather than its entire original design rationale.⁶

Route adds a further distinction. The NUTRIREA-3 investigators describe the preceding NUTRIREA-2 comparison of early enteral and parenteral feeding in ventilated shock as showing no mortality advantage from enteral delivery and more gastrointestinal complications. That earlier finding motivated concern about aggressive feeding in this setting, but NUTRIREA-3 did not randomise route. Its clinician-selected enteral and parenteral pathways therefore cannot establish which route should be preferred for an individual patient.¹

Together, these trials weaken the assumption that more early nutritional substrate necessarily improves meaningful outcomes. They do not establish one optimal dose for all critical illness. Baseline nutritional reserve, illness phase, shock severity, feeding route and the distinction between calories and amino acids remain important differences among the comparisons. The NUTRIREA-3 result is particularly useful because its large achieved contrast and early enrolment make the tested strategy clear. Its limitations are equally specific: a combined intervention, an unblinded recovery endpoint and incomplete evidence about later function.

By the end of 2023, the supported position was therefore selective early restraint rather than a universal restrictive prescription. In comparable ventilated adults with shock, an immediate standard calorie and protein load offered no demonstrated survival advantage and delayed measured readiness for intensive care discharge relative to the low regimen. Decisions beyond the acute week, in established malnutrition, or about protein independent of energy still require evidence addressing those questions directly.¹

Should early calorie and protein delivery be restricted in ventilated adults with shock?

Consider the tested restricted regimen during the acute first week. It hastened readiness for intensive care discharge but did not establish a survival benefit.

References

1. Reignier J, Plantefève G, Mira JP, Argaud L, Asfar P, Aissaoui N, et al.; NUTRIREA-3 Trial Investigators and Clinical Research in Intensive Care and Sepsis (CRICS-TRIGGERSEP) Group. Low versus standard calorie and protein feeding in ventilated adults with shock: a randomised, controlled, multicentre, open-label, parallel-group trial (NUTRIREA-3). *Lancet Respir Med* 2023;11(7):602–612.
2. Rice TW, Wheeler AP, Thompson BT, Steingrub J, Hite RD, Moss M, et al.; National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome Clinical Trials Network. Initial Trophic vs Full Enteral Feeding in Patients With Acute Lung Injury: The EDEN Randomized Trial. *JAMA* 2012;307(8):795–803.
3. Arabi YM, Aldawood AS, Haddad SH, Al-Dorzi HM, Tamim HM, Jones G, et al.; PermiT Trial Group. Permissive Underfeeding or Standard Enteral Feeding in Critically Ill Adults. *N Engl J Med* 2015;372:2398–2408.
4. NUTRIREA-3 Trial Investigators. Supplementary appendix to: Low versus standard calorie and protein feeding in ventilated adults with shock (NUTRIREA-3). *Lancet Respir Med* 2023;11:602–612.
5. Casaer MP, Mesotten D, Hermans G, Wouters PJ, Schetz M, Meyfroidt G, et al. Early versus late parenteral nutrition in critically ill adults. *N Engl J Med* 2011;365:506–517.
6. Chapman M, Peake SL, Bellomo R, Davies A, Deane AM, Horowitz M, et al.; TARGET Investigators and ANZICS Clinical Trials Group. Energy-Dense versus Routine Enteral Nutrition in the Critically Ill. *N Engl J Med* 2018;379:1823–1834.

EFFORT Protein

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.

Introduction

Loss of muscle and impaired physical recovery are central concerns in prolonged critical illness. Providing sufficient protein seems an intuitive response: amino acids are needed for tissue synthesis, and patients often receive less than their prescribed nutritional intake. Yet a larger dose does not necessarily overcome the catabolic response to severe illness. Substrate that cannot be incorporated into useful tissue may instead increase the metabolic work of disposal. Establishing a physiological rationale is therefore only the beginning of the clinical question.¹

The EFFORT Protein investigators described guideline recommendations spanning approximately 1.2–2.0 g/kg/day, with still higher targets for selected populations, despite limited randomised evidence. Their international practice registry also demonstrated a gap between prescribed and delivered protein. That gap could represent a correctable failure of care, but it could equally reflect feeding intolerance, interruptions, illness severity or unrealistic targets. An association between higher intake and better outcomes cannot determine which explanation is responsible.¹

An additional uncertainty concerned patient selection. Malnutrition, frailty, sarcopenia and anticipated prolonged ventilation might identify people with more to gain from nutritional support. They do not necessarily identify patients able to use a large early protein load. Greater nutritional risk may coexist with greater organ dysfunction and reduced metabolic tolerance. A trial enriched for nutritional risk therefore tests an attractive clinical hypothesis, rather than guaranteeing a responsive population.¹

EFFORT Protein converted an established nutritional registry into a pragmatic randomised trial. It asked whether prescribing substantially more protein, with local teams choosing how to deliver it, would improve outcomes in mechanically ventilated adults considered at high nutritional risk. Its international setting offered a test of an implementable prescribing policy across many different intensive care systems.¹

Synopsis

EFFORT Protein was an investigator-initiated, parallel-group, registry-based superiority trial in 85 intensive care units. The abstract and results describe 16 countries, although the methods give a different count. Recruitment occurred between 17 January 2018 and 3 December 2021. There was no core trial funding; UK sites received protein products and limited financial support for local insurance and project management. The suppliers had no role in design, conduct, analysis or publication. The study was registered as NCT03160547.¹

Adults were eligible within 96 hours of intensive care admission if they were expected to require mechanical ventilation for at least another 48 hours and had one or more nutritional risk features. These included body mass index at or below 25 or at or above 35 kg/m², locally assessed moderate or severe malnutrition, Clinical Frailty Scale score at least five, SARC-F score at least four, or anticipated ventilation exceeding four days from screening. SARC-F is a questionnaire concerning strength and physical function. These were alternative entry criteria; established malnutrition was not compulsory.¹

Patients were excluded after more than 96 continuous hours of ventilation, if death or withdrawal of life support was expected within seven days, during pregnancy, or if the treating clinician considered a particular protein dose necessary. Patients needing exclusively parenteral nutrition were excluded where suitable products could not deliver the high target. Consent arrangements varied, including authorised waivers and surrogate consent.¹

A central secure web system allocated patients 1:1, using computer-generated permuted blocks of two, four or eight and stratification by site. Future assignments were concealed. Clinicians prescribing and administering nutrition were unblinded; patients were not informed of allocation. The high group was prescribed at least 2.2 g/kg/day of protein and the usual group no more than 1.2 g/kg/day. Pre-admission dry weight was used, except that patients with body mass index above 30 kg/m² used ideal weight corresponding to 25 kg/m².¹

Assigned feeding began as soon as possible after randomisation within the early eligibility window. Clinicians could combine enteral feeding, parenteral nutrition, intravenous amino acids and enteral protein supplements. Energy was not randomised or fixed; overfeeding was discouraged and other treatment remained at clinicians' discretion. Teams were encouraged to deliver at least 80% of prescribed protein. Assignment continued for up to 28 days, death, or permanent full oral feeding, with continuation after relevant early intensive care readmission.¹

The original primary outcome was 60-day mortality. A planned sample of 4,000 would provide 80% power, at two-sided alpha 0.05, to detect an absolute reduction from 30% to 26%. Pandemic disruption substantially reduced recruitment. On 11 June 2021, after approximately 1,100 patients had been randomised, the steering committee considered exchanging the primary and secondary outcomes. The report states that pooled blinded data informed revised simulations. These suggested that 600 patients per arm would provide 83% power for a combination of reduced hospital mortality and faster discharge among survivors. Recruitment was set to finish in December 2021.¹

The revised primary outcome was time to discharge alive from hospital, measured from randomisation to 60 days after initial intensive care admission. Death was a competing event. Patients dying within that period were classified as never discharged alive even if an earlier discharge had occurred; for readmissions to the index hospital, the last recorded discharge was used. The primary comparison used an unadjusted Fine–Gray subdistribution hazard ratio, with values above one favouring high protein. Site-stratified and covariate-adjusted analyses were also reported. Sixty-day mortality became secondary; other clinical and nutritional outcomes were tertiary. No formal interim efficacy stopping rule was used.¹

Of 4,322 patients assessed, 1,329 were randomised, 659 to high and 670 to usual protein. Fourteen per group were excluded before the analysed intervention cohort, leaving 645 and 656. The exclusions included early departure before intervention, unavailable usable data and withdrawal of permission. Four further patients had no hospital outcome data, so 644 and 653 contributed to the primary time-to-event analysis. Consent withdrawals were censored when withdrawal occurred. Mortality status at 60 days was available for 642 and 648.^{1,2}

Mean age was 57 years in each arm, approximately 60% were men, and more than 80% had medical admissions. Median Sequential Organ Failure Assessment score was nine and median APACHE II score 21. Approximately 40% were receiving vasopressors. The baseline table labelled 163 versus 149 patients as having acute kidney injury at randomisation, and 77 versus 58 were receiving renal replacement therapy. A later 2023 analysis introduced an important qualification to the timing of kidney-injury classification.^{1,3}

Randomisation occurred a median 38.5 versus 39.8 hours after intensive care admission. Enteral nutrition alone was used in 92.6% versus 91.3%. Mean prescribed protein was 2.2 versus 1.2 g/kg/day, while achieved intake was 1.6 versus 0.9 g/kg/day. Mean energy delivery was 14.7 versus 13.2 kcal/kg/day. Protein was assessed during measured ventilated days through study day 28, and energy through day 12. In the high group,

45.0% achieved at least 80% of prescribed protein. Thus the intervention produced substantial separation, but most high-group patients did not meet that adherence threshold.²

The reported cumulative incidence of live hospital discharge by the study's 60-day horizon was 46.1% (95% confidence interval 42.0–50.1) with high protein and 50.2% (46.0–54.3) with usual protein. The unadjusted subdistribution hazard ratio was 0.91 (95% confidence interval 0.77–1.07; $P=0.27$). The fully adjusted estimate was 0.96 (0.80–1.14), based on 1,093 patients with complete covariate data. At 60 days, 222/642 versus 208/648 patients had died, 34.6% versus 32.1%, relative risk 1.08 (95% confidence interval 0.92–1.26).^{1,2}

Mechanical ventilation lasted a median 6.1 days in both groups; other duration outcomes were also similar descriptively. Mean patient-level average urea was higher with high protein, 14.0 versus 11.9 mmol/L, and urea at least 30 mmol/L occurred in 43 versus 20 patients. The subgroup labelled acute kidney injury had a discharge hazard ratio of approximately 0.5 (95% confidence interval 0.4–0.8) and mortality relative risk 1.4 (1.1–1.8). Interaction P values were 0.001 and 0.02, respectively. Higher organ failure scores also suggested less favourable effects, although interaction evidence was weaker. Functional and longer-term performance outcomes were not collected.^{1,2}

Critique

EFFORT Protein provides no evidence that routinely prescribing the higher regimen improves live discharge or survival in its selected population. The point estimates favour usual protein, and the primary interval excludes a large favourable increase in the discharge hazard. However, the mortality interval remains compatible with both modest benefit and appreciable harm. The result does not prove that protein doses are interchangeable, nor that either prescription is optimal for every critically ill patient.¹

The change in primary outcome is a central feature of the appraisal. Reducing a planned 4,000-patient mortality trial to approximately 1,300 patients inevitably sacrifices precision for mortality. The replacement endpoint and revised power calculation asked a different question, with benefit assumed across both survival and recovery. The investigators disclosed the change and reported using blinded pooled data. That is materially different from choosing an outcome after seeing which treatment looked better. Nevertheless, without independent protocol and SAP review, the detailed amendment chronology cannot be verified beyond the principal report. The chapter's conclusion rests on the revised trial as conducted, not on fulfilment of its original mortality ambition.¹

The live-discharge endpoint attempts to prevent early death from appearing to shorten hospital stay. Reclassifying patients who subsequently died and using the final index-hospital discharge also makes it more demanding than simply recording a first discharge. It remains influenced by discharge practices, rehabilitation access and unblinded clinical decisions. The mortality result offers an objective counterweight, but its imprecision limits reassurance. Neither the hazard ratio nor the cumulative incidence should be substituted for a ratio of ordinary median lengths of stay.^{1,2}

The distinction between prescribed and received dose is equally important. The trial tested a policy of prescribing at least 2.2 g/kg/day, but achieved an average near 1.6 g/kg/day. Calling this a trial of everyone receiving 2.2 would misrepresent exposure. Conversely, the 0.7 g/kg/day separation is substantial and makes failure to deliver the entire high target an inadequate explanation for dismissing the result. Escalating further to ensure every patient receives 2.2 would require evidence of benefit and tolerance; it cannot be presumed to rescue the unsuccessful policy.^{1,2}

Energy intake was relatively close, which helps focus the comparison on protein more than trials changing both nutrients dramatically. It was not identical or experimentally controlled, and formulation, route and supplementation were flexible. This pragmatism improves transferability of the prescribing decision while weakening claims about a particular formulation or isolated metabolic mechanism. Achieved-dose analyses would not automatically solve that problem, because patients able to tolerate more nutrition differ from those unable to do so. The randomised allocation remains the most reliable basis for comparing the strategies.¹

The nutritional-risk enrichment was broad. A body mass index of 25 kg/m² or less includes many patients without underweight, and anticipated prolonged ventilation could itself establish eligibility. Frailty and sarcopenia screens were proxy assessments, while malnutrition depended on local assessment. This population should therefore not be described as uniformly severely malnourished or uniformly sarcopenic. Failure of the high regimen in this cohort does not establish an optimum for a narrowly defined nutritional deficiency; equally, a risk label alone does not justify high protein in the absence of demonstrated benefit.¹

The kidney signal deserves attention without overstatement. Its interaction results were stronger than simply finding a significant result inside one subgroup, and the urea difference is biologically compatible with a greater nitrogen disposal burden. Yet several subgroups and two outcomes were examined without multiplicity correction. Kidney injury and higher organ failure scores also overlap clinically, so they do not constitute

independent replications. The weaker organ-failure interactions should not be presented with the same certainty as the kidney finding.^{1,2}

A further limitation emerged within the book's evidence window. The October 2023 renal analysis defined injury within seven days of intensive care admission and acknowledged classification using post-randomisation data. This qualifies the principal report's baseline wording. Conditioning on an event occurring after allocation can undermine the protection of randomisation; the subgroup cannot be treated as a clean pretreatment selector. The signal supports caution, but does not prove a dose rule for every patient with kidney injury.³

Higher urea is evidence of a biochemical effect, not proof of the mechanism causing worse clinical outcomes. Intake, endogenous breakdown, renal clearance and renal replacement therapy all influence urea. The supplied metabolic table summarises patient-level averages and uses varying measurement denominators. Neither an average concentration nor a threshold count can establish that excessive amino acid breakdown mediated death. Mechanistic explanations should remain hypotheses rather than reasons to regard the subgroup result as settled.^{1,2}

Post-randomisation exclusions also require acknowledgement. The primary cohort omitted 28 patients, including patients leaving before receiving the intervention. Removing early deaths or rapid recoveries can introduce selection even when the number excluded is balanced. The small proportion makes a large distortion less likely, but this remains a modified intention-to-treat analysis. Further losses affect endpoint denominators, and the fully adjusted model excludes additional patients with missing covariates. Agreement across analyses is reassuring, although adjustment is not a substitute for retaining the full randomised population.^{1,2}

The absence of functional outcomes limits the trial's ability to address its most attractive biological rationale. More protein might conceivably affect muscle or rehabilitation without changing short-term discharge or mortality; EFFORT did not measure that possibility. It also maintained one assigned target across different illness phases rather than randomising early and later protein strategies separately. A later rehabilitation intervention with exercise would be a different treatment, not an explanation that can be assumed to overturn the present findings.¹

Finally, geographical breadth does not ensure equal evidence for every diagnosis. The cohort was predominantly medical, with relatively few surgical, burn or trauma patients. Clinician equipoise exclusions also remove patients thought to require a specific dose. The defensible practice implication is to avoid routine escalation to the tested high

prescription in comparable ventilated adults, particularly when organ dysfunction raises concern. This leaves room for individual reassessment and separate evidence in populations poorly represented by the trial.¹

Body of Evidence

The earlier literature presented a tension between nutritional physiology and patient-centred outcomes. Observational associations could make more protein appear attractive, while small randomised trials often struggled to achieve separation or estimate clinical effects precisely. EFFORT's own review of prior trials found an inconsistent and limited evidence base. Its contribution was therefore not another demonstration that protein can be delivered, but a much larger test of whether a higher prescribing policy improves recovery in patients selected for nutritional risk.¹

A blinded feasibility trial by Chapple and colleagues illustrates the distinction. It randomised 120 patients across six intensive care units, with data available for 116. Formula containing more protein achieved approximately 1.52 versus 0.99 g/kg ideal body weight/day with similar energy delivery. Ninety-day mortality was 26% versus 27%, with a wide confidence interval for the difference. This established that a masked formula comparison with meaningful protein separation was feasible; it did not establish clinical equivalence or a mortality benefit.⁴

Kidney function had already emerged as a possible modifier. A 2018 post hoc analysis of a 474-patient intravenous amino acid trial reported an interaction with baseline kidney status. It suggested lower mortality in patients with normal kidney function, but did not demonstrate a mortality difference in the subgroup with dysfunction or risk of worsening injury. This exploratory finding is not identical to proof that high protein harms patients with renal impairment. It provides a reason to investigate heterogeneity, rather than an independent confirmation of EFFORT's adverse subgroup estimate.⁵

The 2023 EFFORT renal report explored the same 312 patients, rather than providing a new randomised replication. It found no significant modification by its further subgroups and no significant difference in new renal replacement therapy. Apparent attenuation of harm during dialysis therefore did not establish a protected group in whom high protein should be prescribed.³

NUTRIREA-3, also published in 2023, addressed a complementary question in ventilated shock. It combined marked calorie restriction with low protein during the acute first week. Among 3,036 analysed patients, 90-day mortality was 41.3% versus 42.8%, while readiness for intensive care discharge occurred earlier with the restricted strategy. Its

actual protein intake, 0.2 versus 0.9 g/kg/day, differed from EFFORT's 1.6 versus 0.9 g/kg/day comparison. Calories and fluid also separated markedly, so it cannot identify an independent benefit of protein restriction.⁶

These two large trials should not be placed on a single dose-response curve. They enrolled different populations, used different weight rules, began treatment within different windows and maintained assignment for different durations. EFFORT permitted a higher prescription for up to 28 days; NUTRIREA-3 explicitly increased both groups' targets after the early acute phase. The shared absence of demonstrated benefit from the more intensive regimen challenges automatic escalation, but does not locate one optimum protein dose across all phases of illness.^{1,6}

The practical evidence gap is consequently narrower and more useful than a general claim that nutrition does not work. It concerns who benefits from additional protein, when that benefit can occur, which functional outcomes matter, and whether renal or other organ dysfunction alters tolerance. Trials testing these questions need clear timing, adequate exposure separation and outcomes extending beyond the ability to leave hospital. They also need to distinguish pretreatment patient characteristics from events that arise after randomisation.

By the end of 2023, EFFORT supported a more restrained approach to routine high-protein prescribing in mechanically ventilated adults. The usual-dose arm performed at least as well on the observed clinical outcomes, but the study was not a non-inferiority trial and did not certify a universal safe minimum or maximum. Its most reliable contribution is the failure of the tested high prescription to improve recovery, accompanied by a clinically concerning, methodologically qualified signal of harm in patients with kidney injury.^{1,2}

Should nutritionally high-risk ventilated adults routinely receive a protein prescription of at least 2.2 g/kg/day?

No. The tested higher prescription did not improve recovery and raised concern about harm in patients with kidney injury.

References

1. 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.

2. EFFORT Protein Trial team. Supplementary appendix to: The effect of higher protein dosing in critically ill patients with high nutritional risk (EFFORT Protein). *Lancet* 2023;401:568–576.
3. Stoppe C, Patel JJ, Zarbock A, Lee ZY, Rice TW, Mafrici B, et al. The impact of higher protein dosing on outcomes in critically ill patients with acute kidney injury: a post hoc analysis of the EFFORT protein trial. *Crit Care* 2023;27:399.
4. Chapple LS, Summers MJ, Bellomo R, Chapman MJ, Davies AR, Ferrie S, et al.; TARGET Investigator Collaborative and ANZICS Clinical Trials Group. 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.
5. Zhu R, Allingstrup MJ, Perner A, Doig GS; Nephro-Protective Trial Investigators Group. The Effect of IV Amino Acid Supplementation on Mortality in ICU Patients May Be Dependent on Kidney Function: Post Hoc Subgroup Analyses of a Multicenter Randomized Trial. *Crit Care Med* 2018;46(8):1293–1301.
6. Reignier J, Plantefeve G, Mira JP, Argaud L, Asfar P, Aissaoui N, et al.; NUTRIREA-3 Trial Investigators and Clinical Research in Intensive Care and Sepsis (CRICS-TRIGGERSEP) Group. Low versus standard calorie and protein feeding in ventilated adults with shock: a randomised, controlled, multicentre, open-label, parallel-group trial (NUTRIREA-3). *Lancet Respir Med* 2023;11(7):602–612.



Liver Trials

AntibioCor

Louvet A, Labreuche J, Dao T, Thévenot T, Oberti F, Bureau C, et al. Effect of Prophylactic Antibiotics on Mortality in Severe Alcohol-Related Hepatitis: A Randomized Clinical Trial. JAMA 2023;329(18):1558–1566.

Introduction

Severe alcohol-related hepatitis combines substantial hepatic inflammation with impaired host defence. Patients may deteriorate through liver failure, kidney dysfunction, bleeding or infection, and these processes often coexist. Corticosteroid treatment introduces a further tension: suppressing inflammation may help selected patients, while increasing susceptibility to serious infection. Preventing that complication is an attractive strategy, but only if the benefit outweighs the adverse consequences of exposing patients without active infection to antibiotics.^{1,2}

The rationale for prophylaxis extends beyond preventing a discrete episode of sepsis. The AntibioCor investigators describe experimental evidence linking gut-derived bacterial products to hepatic inflammation and suggesting that antibiotic treatment could reduce alcohol-related liver injury. This proposed mechanism would make antibiotics an adjunct to disease treatment as well as infection prevention. It remained uncertain whether such effects would translate into improved liver function or survival in hospitalised patients.¹

The distinction between association and intervention is particularly important here. Infection during corticosteroid treatment identifies patients with poor outcomes, but infection may be a cause of deterioration, a consequence of severe disease, or both. Even effective prevention of some bacterial infections may not reverse the underlying hepatic injury. Antibiotics also have incomplete microbial coverage and may select resistant organisms, so the direction of their overall clinical effect cannot be inferred from infection risk alone.^{1,2}

AntibioCor tested a practical version of this hypothesis: one month of oral amoxicillin–clavulanate added to prednisolone in biopsy-confirmed severe alcohol-related hepatitis. Its primary objective was improved survival at 60 days. The trial therefore required more than a reduction in recorded infection to establish the clinical benefit for which routine prophylaxis was being proposed.¹

Synopsis

AntibioCor was a multicentre, randomised, double-blind, placebo-controlled superiority trial in France and Belgium. The study described 25 participating centres, with recruitment reported from 24. Enrolment ran from 13 June 2015 to 24 May 2019, and final follow-up occurred on 19 November 2019. French Ministry of Health funding supported the trial through a hospital clinical research programme grant to Lille; the funders had no role in trial conduct, analysis or reporting. Registration was NCT02281929.¹

Patients aged 18–75 years required recent jaundice, a compatible clinical diagnosis and histological confirmation of alcohol-related hepatitis, usually through transjugular biopsy. Reported alcohol consumption had to be at least 40 g/day for women or 50 g/day for men. Both a Maddrey discriminant function of at least 32 and a Model for End-stage Liver Disease score of at least 21 were required. These thresholds selected severe hepatic disease rather than an unselected population with alcohol-related liver injury.¹

Exclusions included allergy to the study antibiotic, type 1 hepatorenal syndrome before treatment, uncontrolled gastrointestinal bleeding, untreated infection, hepatitis C viraemia and hepatitis B surface antigen positivity. If bacterial sepsis was present at admission, it had to be treated successfully, followed by at least seven days without antibiotics before trial entry. Patients therefore entered a prophylaxis comparison rather than a trial withholding treatment for active infection. Written consent was obtained from patients or relatives when severe encephalopathy prevented patient consent.^{1,3}

An independent statistician supplied the allocation list. Central contact with the sponsor returned the randomisation number and numbered treatment box. Allocation was 1:1, in blocks of four, stratified by centre. Both patients and investigators were masked through placebo treatment. All patients were prescribed oral prednisolone 40 mg/day for 30 days. The antibiotic arm received amoxicillin 1 g plus clavulanate 125 mg three times daily for 30 days; controls received matching oral-suspension placebo packets at the same frequency.¹

The study did not recommend discontinuing prednisolone for a Lille score of 0.45 or greater at day seven, the threshold identifying poor biochemical response. Participants attended weekly during the first four weeks, then at days 45 and 60 and at three and six months. Diaries, unused medication and interviews informed investigator assessments of adherence, classified as good or poor without a formal numerical definition. Treatment could be interrupted as clinical circumstances required.¹

The primary outcome was all-cause mortality from randomisation to day 60. Death rates were estimated with Kaplan–Meier methods and compared using the log-rank test and an unadjusted Cox hazard ratio. Patients undergoing transplantation, or lacking further vital-status information, were censored. Secondary outcomes included later mortality, infection and hepatorenal syndrome by day 60, day-seven Lille response and improvement in MELD score by day 60. The report describes differences between the protocol and analysis-plan lists of secondary outcomes; the principal 60-day mortality endpoint was unchanged.¹

Infections were diagnosed by local investigators using clinical and biological findings, without a prespecified uniform definition or formal adjudication. Their cumulative incidence and that of hepatorenal syndrome treated death as a competing event, with Fine–Gray subhazard ratios. After the infection analysis failed the proportional-hazards assumption, a post hoc analysis separated days zero to 30 from days 31 to 60. Other post hoc analyses included death or transplantation as a composite and alcohol relapse. Missing values were not imputed, and secondary analyses were exploratory because of multiple comparisons.¹

The planned sample was 280, allowing 5% dropout and 80% power at two-sided alpha 0.05. The assumed control mortality was 27%, and the target was an absolute reduction of 14 percentage points. This large effect was derived from a previous difference in prognosis between MELD strata, rather than a demonstrated antibiotic treatment effect. The supplementary calculation anticipated approximately 54 deaths.^{1,3}

Of 297 patients invited for assessment, 292 were randomised, 145 to antibiotics and 147 to placebo. Three and five respectively were excluded because consent was absent or withdrawn, leaving 142 in each efficacy group. Thirteen patients received the opposite treatment through centre errors, and two received no assigned intervention. Efficacy analysis retained allocation; safety analysis used treatment received. A sensitivity analysis excluded major deviations, leaving 132 and 134 patients.^{1,3}

The analysed patients were approximately 53 years old, about 72% were men and mean MELD scores were 24.9 and 25.8. Median Maddrey scores were 64.1 and 69.7. Encephalopathy was present in 12% of each group, and median creatinine was 61.7 versus 70.5 $\mu\text{mol/L}$. Previously treated infection was recorded in 31 versus 26 patients. Investigator-rated adherence was generally high, but by day 28 good study-drug adherence was recorded in 86.3% versus 78.2%. Temporary or permanent interruption occurred in 19 antibiotic-treated versus 36 placebo-treated patients.^{1,3}

By day 60, 24 versus 31 deaths had occurred among the 142 patients assigned to each group. The principal results table gives Kaplan–Meier mortality estimates of 17.3% versus 21.9%, an absolute difference of –4.7 percentage points (95% confidence interval –14.0 to 4.7), and hazard ratio 0.77 (95% confidence interval 0.45–1.31; $P=0.33$). Mortality at 90 days was 21.0% versus 26.3%, and at 180 days 31.3% versus 34.4%, without a statistically significant difference at either time.¹

The cumulative incidence of infection at day 60 was 29.7% versus 41.5%, with 42 versus 59 events entering that analysis. The absolute difference was –11.8 percentage points (95% confidence interval –23.0 to –0.7), subhazard ratio 0.62 (95% confidence interval 0.41–0.91; $P=0.02$). In the post hoc time-period analysis, the difference arose during the treatment month: 13.4% versus 28.9%, subhazard ratio 0.41 (0.24–0.70). During days 31–60, incidences were similar, with subhazard ratio 1.06 (0.57–1.96).¹

Hepatorenal syndrome occurred with estimated incidences of 7.7% versus 8.5%, subhazard ratio 0.91 (95% confidence interval 0.40–2.05). Lille response was observed in 80/141 versus 76/138 assessed patients, and a MELD score below 17 at day 60 in 48/100 versus 50/91 assessed survivors without transplantation. Neither comparison demonstrated benefit. The day-60 death-or-transplant estimate was 19.8% versus 23.2%, hazard ratio 0.84 (0.50–1.40).^{1,3}

In the safety population, 68/141 antibiotic-treated and 77/143 placebo-treated patients experienced serious adverse events through day 60, comprising 105 versus 125 events. No drug-induced liver injury was recorded. Among 185 patients with rectal swabs, multidrug-resistant bacterial carriage was found in 29/91 versus 8/94. The subgroup with previously treated infection had a favourable but imprecise mortality estimate, hazard ratio 0.40 (95% confidence interval 0.13–1.16); the interaction P value was 0.17.^{1,3}

Critique

The principal result is a failure to establish survival benefit, despite fewer recorded infections. This does not establish equivalence. The mortality confidence interval permits a large benefit and some harm, and the observed difference is compatible with a clinically useful effect that the study could not resolve. However, routine prophylaxis requires positive evidence of overall benefit, not simply an interval that leaves benefit possible. The infection finding alone does not meet the trial's original objective.¹

The power calculation explains much of this uncertainty. A reduction in mortality from 27% to approximately 13% is ambitious for an adjunct intended partly to prevent one category of complication. The actual placebo mortality was lower than assumed,

although the total observed deaths were close to the supplementary event target. It would therefore be inaccurate to attribute the neutral finding only to too few events. The more fundamental limitation is the large treatment effect the trial was designed to detect. Its precision is insufficient for a definitive claim that smaller survival gains do not exist.^{1,3}

The causal contrast is otherwise well protected. Concealed central allocation, placebo masking and an objective primary endpoint limit several important biases. Biopsy confirmation strengthens diagnostic specificity. The practical administration errors weakened separation but were retained in the allocated efficacy groups, avoiding the bias of simply reassigning patients according to what they received. The analysis excluding major deviations gave a similar mortality estimate. That consistency is reassuring, although a per-protocol sensitivity analysis cannot recreate the original randomisation after exclusions.^{1,3}

Transplantation complicates interpretation of survival. Censoring at transplant assumes that subsequent unobserved survival can be handled like other censoring, even though transplant is a major rescue intervention selected for patients with a particular prognosis. Four versus two patients underwent transplant by day 60, and five versus seven over the full study. The post hoc death-or-transplant analysis was directionally similar and did not demonstrate benefit, which reduces concern that the main conclusion depends entirely on censoring. It remains a different estimand, and should be described as such.^{1,3}

The infection endpoint is less secure than death ascertainment. Local clinical diagnosis without a uniform prespecified definition can combine events of different certainty and severity. Masking reduces differential diagnostic behaviour, but does not eliminate variation between centres or distinguish microbiological prevention from suppression of signs and culture yield. The latter is a possible interpretation, not an effect established by this trial. The results support fewer diagnosed infections under this regimen, with appropriate caution about precisely what was prevented.¹

The temporal pattern is plausible: benefit was concentrated while antibiotics were being administered and was not apparent after cessation. However, splitting follow-up into two periods was undertaken after checking the model assumption, so those estimates are exploratory. The second period also concerns patients remaining at risk, a population shaped by earlier infection and survival. It cannot establish that prophylaxis merely postpones every infection or that extending antibiotics would improve outcomes. Either claim would require a different randomised comparison.¹

The absence of improvement in Lille response, MELD or hepatorenal syndrome weakens the proposed disease-modifying mechanism. These outcomes were not uniformly available for the entire randomised cohort. In particular, day-60 MELD comparisons were among assessed survivors without transplantation; they should not be interpreted as the probability that any randomised patient achieved recovery. Their neutrality is compatible with no substantial hepatic benefit, but neither proves that gut-derived bacterial products are irrelevant nor identifies the biological reason survival failed to improve.^{1,3}

Corticosteroid management limits transferability. All patients were intended to receive 30 days of prednisolone without a specified Lille-based stopping recommendation. Continuing immunosuppression in nonresponders could alter both infection risk and the opportunity for prophylaxis to help. The trial therefore appraises antibiotics within that steroid strategy. It does not establish their value when steroid exposure is actively shortened according to response. Nor does it justify changing corticosteroid treatment solely to make antibiotic prophylaxis appear more attractive.^{1,2}

Antimicrobial ecology is a substantive counterweight to the infection result. The marked difference in resistant bacterial carriage among sampled patients is concerning, particularly in a population vulnerable to subsequent infection. Sampling covered only 185 of 284 analysed patients, and the supplied report does not fully clarify baseline carriage, serial sampling or why other patients were not sampled. The finding therefore cannot quantify new acquisition for the complete cohort or prove a corresponding increase in resistant clinical infection. It does prevent a reassuring narrative based solely on fewer recorded infections and the absence of observed drug-induced liver injury.¹

Adherence and adverse-event reporting also need careful denominators. Good adherence was a judgement rather than a quantified proportion of doses taken, and interruptions were more frequent with placebo. Some interruptions may reflect illness or infection arising during treatment, so they cannot simply be treated as unrelated noncompliance. Safety tables classify patients by exposure and distinguish patients from events; their infection categories are not interchangeable with the allocated efficacy analysis. Small numbers of *Clostridioides difficile* or fungal events do not establish the absence of these harms.^{1,3}

The prior-infection subgroup is insufficient to create an indication for routine continuation of antibiotics. Its point estimate is favourable, but the interval crosses the null and the interaction test does not establish a different treatment effect. More importantly, entry required a seven-day antibiotic-free interval after cure. That design differs from continuing therapeutic antibiotics across the initiation of corticosteroids. The accompanying editorial's suggestion that overlap might be considered is a clinical

interpretation informed by other observations, not a randomised result demonstrated by AntibioCor.^{1,2,3}

Applicability is strongest to patients with biopsy-confirmed severe hepatitis who can enter after treatment of active infection and who do not already have type 1 hepatorenal syndrome or uncontrolled bleeding. Intensive care admission or organ support was not an entry requirement. The trial should not be presented as a study of antibiotics for septic shock or for all forms of decompensated cirrhosis. It also tests one oral antibiotic regimen and duration; its neutral survival finding neither endorses other prophylactic regimens nor removes separate indications for therapeutic antibiotics.^{1,3}

Body of Evidence

The preceding corticosteroid evidence explains the attraction of antibiotic protection. In the 1,103-patient STOPAH trial, the unadjusted odds ratio for 28-day mortality with prednisolone was 0.72 (95% confidence interval 0.52–1.01); adjustment for baseline severity strengthened the estimate. There was no significant later mortality advantage at 90 days or one year. Serious infections occurred in 13% with prednisolone versus 7% without it. The combination of a possible early anti-inflammatory benefit and increased serious infection made an adjunctive prevention strategy worth testing.⁴

The supplied editorial also reviews analyses linking infection during corticosteroid treatment to poor prognosis. It describes a STOPAH observation suggesting better outcomes when treatment of baseline infection overlapped with steroid initiation. That antibiotic exposure was not itself randomised. Patients whose treatment continued may have differed in infection recognition, treatment response or other aspects of care. Such findings can generate the hypothesis tested by AntibioCor, but cannot substitute for a positive prophylaxis trial or resolve the uncertainty of its small prior-infection subgroup.²

Direct antibiotic evidence before AntibioCor was limited. A 32-patient randomised study by Kimer and colleagues compared standard treatment with or without rifaximin, focusing on inflammatory and metabolic markers. It found no significant differences in the measured amino acid or inflammatory responses at 28 and 90 days. Four control patients and five rifaximin patients died. The small size and mechanistic primary outcomes preclude a reliable survival conclusion, but the trial did not provide convincing clinical confirmation of the proposed gut–liver mechanism.⁵

In adjacent liver disease, the AntibioCor investigators cite a 291-patient trial of norfloxacin in advanced Child–Pugh C cirrhosis that reduced infections without demonstrating an overall survival benefit. They also describe a small rifaximin

case-control study in alcohol-related hepatitis with fewer infections but no survival advantage. These accounts are consistent with a recurring distinction between infection prevention and mortality, although the populations, drugs and designs differ. They should be regarded as contextual support rather than pooled replications of the amoxicillin–clavulanate comparison.¹

Forrest and Bernal interpret AntibioCor as evidence against giving prophylactic antibiotics to every patient receiving corticosteroids for severe alcohol-related hepatitis. They also propose that failure to improve liver function despite fewer infections is compatible with infection being a consequence of severe disease or its treatment. That is one plausible explanation, but the trial cannot separate it from limited power, incomplete microbial coverage or competing pathways to death. The editorial argument is useful when retained as interpretation rather than promoted to a demonstrated causal mechanism.²

AntibioCor materially improves this evidence base by randomising a defined antibiotic regimen against placebo in a well-characterised population and following survival beyond treatment. Its result is especially relevant to the temptation to convert a high baseline infection risk into universal prophylaxis. A high risk identifies the potential burden of a complication; it does not by itself establish that preventing some episodes improves the outcome that matters most to patients.

The unresolved research questions concern better selection, the timing of treatment relative to infection and corticosteroids, and whether alternative microbial strategies can improve patient-centred outcomes without unacceptable resistance. Those questions should preserve the distinction between prophylaxis and treatment of an established infection. A trial of a different strategy would also need standardised infection ascertainment, complete resistance surveillance and enough precision to assess a realistic survival effect.

By the end of 2023, the evidence did not support routine one-month amoxicillin–clavulanate prophylaxis to improve survival in this setting. Prompt recognition and treatment of active infection remain essential parts of care, while decisions about steroid continuation and other liver-related indications require their own assessment. AntibioCor narrows one prescribing decision; it does not reduce the importance of infection in the clinical management of severe alcohol-related hepatitis.^{1,2}

Should amoxicillin–clavulanate routinely accompany prednisolone in severe alcohol-related hepatitis?

No. It reduced recorded infections but did not establish a survival benefit, and resistant bacterial carriage was more frequent among sampled patients.

References

1. Louvet A, Labreuche J, Dao T, Thévenot T, Oberti F, Bureau C, et al. Effect of Prophylactic Antibiotics on Mortality in Severe Alcohol-Related Hepatitis: A Randomized Clinical Trial. *JAMA* 2023;329(18):1558–1566.
2. Forrest E, Bernal W. The Role of Prophylactic Antibiotics for Patients With Severe Alcohol-Related Hepatitis. *JAMA* 2023;329(18):1552–1553.
3. Louvet A, Labreuche J, Dao T, Thévenot T, Oberti F, Bureau C, et al. Supplemental Online Content to: Effect of Prophylactic Antibiotics on Mortality in Severe Alcohol-Related Hepatitis. *JAMA* 2023;329:1558–1566.
4. Thursz M, Forrest E, Roderick P, Day C, Austin A, O’Grady J, et al. The clinical effectiveness and cost-effectiveness of STeroids Or Pentoxifylline for Alcoholic Hepatitis (STOPAH): a 2 × 2 factorial randomised controlled trial. *Health Technol Assess* 2015;19(102):1–104.
5. Kimer N, Meldgaard M, Hamberg O, Kronborg TM, Lund AM, Møller HJ, et al. The impact of rifaximin on inflammation and metabolism in alcoholic hepatitis: A randomized clinical trial. *PLoS One* 2022;17(3):e0264278.



Haematology Trials

PACER

van Baarle FLF, van de Weerdt EK, van der Velden WJFM, Ruiterkamp RA, Tuinman PR, Ypma PF, et al. Platelet Transfusion before CVC Placement in Patients with Thrombocytopenia. *N Engl J Med* 2023;388(21):1956–1965.

Introduction

Central venous access creates an immediate haemostatic challenge in patients whose platelet count is already low. Giving platelets before puncturing a large vein is biologically plausible, but exposes patients to transfusion reactions, consumes a scarce resource and may delay access. The clinical decision also depends on the site, catheter, operator and reason for thrombocytopenia. A platelet count cannot describe all these influences. Consequently, a threshold that appears reasonable for an uncomplicated internal jugular insertion may not apply to a tunnelled catheter or a less compressible site.

Ultrasound guidance and experienced operators have reduced procedural trauma. Observational reports of few serious bleeds therefore encouraged a more restrictive approach to prophylactic transfusion. However, these series could not determine what would have happened to comparable patients if platelets had been withheld. Clinicians may preferentially transfuse the patients at greatest risk, while retrospectively recorded outcomes may miss bleeding controlled by prolonged compression. Before PACER, guidance varied and the investigators described an unresolved evidence gap for procedural platelet transfusion in patients with counts between 10 and $50 \times 10^9/L$.¹

PACER directly tested omission of a prophylactic platelet unit before ultrasound-guided central venous catheter placement. Its importance lies in addressing a common bedside decision with random allocation and prospective bleeding assessment. Its interpretation requires equal care: the unit of randomisation was a catheter procedure, the primary outcome included bleeding managed with relatively minor interventions, and the study was designed to establish non-inferiority. Failure to establish that margin, evidence of increased bleeding, and proof of increased life-threatening harm are distinct conclusions. Keeping them separate determines how far the result should change practice.

Synopsis

PACER was a multicentre randomised non-inferiority trial conducted in haematology wards and ICUs at ten Dutch hospitals, seven academic and three general. Recruitment ran from February 2016 to March 2022. Adults older than 18 years who needed a central venous catheter and had a platelet count of $10\text{--}50 \times 10^9/\text{L}$ were eligible. The principal report required an anticipated catheter duration of at least 24 hours. Important exclusions included coagulation factor deficiency, a bleeding diathesis and therapeutic anticoagulation, although the supplementary criteria allowed continuous heparin stopped at least one hour before insertion. Single antiplatelet treatment was permitted. Correction of an initially lower platelet count or excessive INR before randomisation was allowed if a subsequent measurement met eligibility.^{1,2}

The upper INR threshold increased from 1.5 to 3.0 after approximately two-thirds of the trial had been completed. This broadened eligibility during recruitment, although the median baseline INR was 1.1 in both groups. Haematology patients provided prospective consent; deferred consent was used in intensive care. A patient could contribute more than one catheter placement, provided no randomisation had occurred during the preceding 24 hours.

Catheter episodes were assigned equally to one unit of platelet concentrate before insertion or no prophylactic transfusion, with stratification by centre and regular versus large-bore dialysis catheter. Standard products were pooled buffy-coat platelet concentrates; matched or irradiated products could be used when indicated. Patients and treating clinicians knew allocation. Operators were blinded when possible, rather than uniformly. In the per-protocol population, operator blinding was recorded for 143 of 188 transfusion procedures and 142 of 185 no-transfusion procedures.²

Ultrasound guidance was mandatory and operators had performed at least 50 previous central venous catheter insertions. Catheters were generally placed approximately one hour after randomisation. Internal jugular, subclavian and femoral sites, regular and dialysis catheters, and tunnelled and non-tunnelled devices were represented. Rescue treatment remained available. The supplementary methods encouraged inspection, compression, consideration of a skin suture, platelet transfusion and, if required, radiological or surgical intervention. Assignment to no prophylaxis therefore did not mean that subsequent platelet transfusion was prohibited.

The primary outcome was catheter-related bleeding of grade 2–4 within 24 hours. Trained staff assessed bleeding immediately after insertion and at one and 24 hours. Grade 2 included bleeding requiring a minor intervention, such as more than 20 minutes of compression. Grade 3 included red-cell transfusion without haemodynamic instability

or a radiological or elective surgical intervention. Grade 4 represented severe instability or fatal bleeding. Grade 1 covered lesser bleeding, including compression lasting less than 20 minutes. Major bleeding of grade 3–4 was the key secondary outcome; other measures included haematoma, transfusions, transfusion reactions, lung injury, length of stay, mortality and costs.¹

The sample-size calculation assumed a 1% primary-event risk with transfusion and an acceptable absolute excess of 2.5 percentage points without transfusion. This was translated into a relative-risk non-inferiority margin of 3.5. A target of 392 catheter episodes provided 80% power at a one-sided significance level of 0.05. The primary per-protocol analysis used a two-sided 90% confidence interval, equivalent to a one-sided 95% bound for the non-inferiority decision. Intention-to-treat analysis was also performed. Poisson mixed models included treatment and catheter type as fixed effects and centre as a random effect. Secondary intervals were 95% and were not adjusted for multiple comparisons.

Of 411 randomised episodes, 18 were excluded because deferred consent was not obtained or consent was withdrawn. The intention-to-treat population comprised 393 episodes in 358 patients, allocated 197 to transfusion and 196 to no transfusion. Another 20 episodes were excluded from the per-protocol population for eligibility or treatment deviations, leaving 373 episodes in 338 patients: 188 with transfusion and 185 without. Around 57% of these procedures occurred on haematology wards and 43% in ICUs. Median platelet count was $30 \times 10^9/\text{L}$ in each group. Six episodes crossed from assigned transfusion to no transfusion and eight crossed in the opposite direction. Platelets had also been given during the six hours before randomisation in 8.5% and 10.3% of per-protocol episodes, respectively, so the comparison was between procedural prophylaxis strategies against a background of ongoing clinical transfusion practice. Approximately half used the internal jugular vein, 38% the subclavian vein and 12% the femoral vein; about one in ten catheters was tunnelled.^{1,2}

The primary outcome was incompletely recorded in 27 per-protocol episodes, or 7.2%, and was analysed using multiple imputation. Grade 2–4 bleeding occurred in an estimated 9 of 188 episodes with transfusion, 4.8%, and 22 of 185 without transfusion, 11.9%. The relative risk for no transfusion versus transfusion was 2.45 (90% CI 1.27–4.70). The reported absolute risk difference was 7.1 percentage points (90% CI 1.3–17.8). The upper relative-risk bound exceeded 3.5, so non-inferiority was not established. The intention-to-treat estimate was similar, RR 2.53 (90% CI 1.32–4.86).^{1,2}

Grade 3–4 bleeding occurred in four transfusion episodes, 2.1%, and nine no-transfusion episodes, 4.9%, giving RR 2.43 (95% CI 0.75–7.93). All 13 were grade 3 bleeds requiring

red cells; there were no grade 4 events. Grade 1 bleeding and haematoma were numerically more frequent without prophylaxis, but their per-protocol confidence intervals included no difference. Platelet counts at one hour were higher with prophylaxis, with medians of 54 versus $26 \times 10^9/L$. Platelet use during the subsequent 24 hours was higher in the no-prophylaxis group; this measure excluded the assigned preprocedural unit. Hospital stay and mortality did not show a clear treatment effect.

The adverse-event summary included the 13 major bleeds, one episode of transfusion-related acute lung injury and one transient Horner syndrome. The detailed reaction and cost tables used additional categories. Withholding prophylaxis reduced estimated transfusion- and bleeding-related expenditure by US\$410 per catheter placement (95% CI 285–545), despite greater subsequent transfusion expenditure. The investigators concluded that omission failed the non-inferiority criterion and increased catheter-related bleeding. Their suggestions for tailoring transfusion to platelet count, department or catheter characteristics extended beyond the overall randomised comparison.^{1,2}

Critique

The non-inferiority margin nevertheless deserves scrutiny. An acceptable absolute increase of 2.5 percentage points was linked to an anticipated 1% control-event rate, producing a relative-risk margin of 3.5. The observed control rate was 4.8%. A fixed relative margin at that higher background risk would allow a much larger absolute increase than originally envisaged. The trial failed even this relative criterion. More importantly, the primary interval also lay above one, supporting an increase in the composite bleeding outcome on the reported scale. This is stronger evidence than an inconclusive non-inferiority test alone, but it does not establish a similarly precise effect on major bleeding.¹

PACER provides unusually direct evidence for a procedural transfusion decision. Random allocation addresses confounding by indication, which is particularly troublesome when clinicians give platelets to the patients they consider most likely to bleed. Mandatory ultrasound guidance and experienced operators make the comparison relevant to competent contemporary practice. Including haematology and intensive care settings, several insertion sites and dialysis catheters broadens its clinical scope. The result cannot readily be dismissed as a consequence of routinely poor insertion technique.

The outcome combines events with substantially different consequences. Prolonged compression is inconvenient, consumes staff time and matters to patients, yet is not equivalent to haemodynamic collapse. There were no grade 4 bleeds, and the

major-bleeding interval extended from a possible reduction to nearly an eightfold increase without prophylaxis. The data therefore favour prophylaxis for overall catheter-related bleeding while leaving the magnitude of its effect on severe harm uncertain. Neither a claim that omission is harmless nor a claim that prophylaxis prevents fatal bleeding is justified.

Blinding was incomplete at several levels. Approximately three-quarters of operators were blinded, but the patient and clinical team were not. Knowledge that platelets had been withheld might influence the duration of compression, the decision to transfuse red cells, or the threshold for recording a minor intervention. Operator blinding addresses insertion behaviour more directly than subsequent outcome assessment. The post hoc analysis among blinded operators remained directionally consistent with the main result, but cannot remove bias arising after the procedure. The in-scope reports do not provide enough detail to independently reconstruct sequence generation and allocation concealment, and these safeguards should not be assumed.²

Missing primary outcomes are material when only about 31 events underpin the result. Twenty-seven per-protocol episodes lacked complete grade 2–4 ascertainment. Investigators retrospectively reviewed records on the assumption that a major bleed would have been documented, but that a lesser bleed might not have been. This explains why major bleeding had no missing values while the broader primary outcome did. Multiple imputation depends on the adequacy of measured predictors and cannot prove that undocumented minor bleeding was comparable between groups. Reassuringly, complete-case analysis gave RR 2.31 (90% CI 1.19–4.49), again failing non-inferiority and favouring prophylaxis. Agreement reduces, without eliminating, the concern.²

The denominator requires particular care. The principal comparison was between catheter episodes, not 373 independent people. Some patients had up to four study catheters, potentially creating correlated bleeding susceptibility and previous transfusion exposure. The published principal model describes centre-level clustering rather than a patient-level random effect. Analyses retaining only the first or a randomly selected catheter per patient were therefore useful and remained consistent with the main result. Their estimates were less precise. These checks support the direction of the finding but do not make repeated procedures equivalent to independent patients.

The per-protocol emphasis is conventional for non-inferiority because crossovers can make treatments appear artificially similar. However, exclusions after allocation can also impair randomisation. PACER strengthens its interpretation by showing concordant intention-to-treat results, although that population itself excluded 18 randomised episodes for consent reasons. The separate flow counts should remain visible. There is

no basis for calling this a complete analysis of every randomised episode or for treating the absence of loss to follow-up as absence of missing outcome data.^{1,2}

Subgroup findings are insufficient to create a validated transfusion algorithm. The ICU estimate was RR 1.36 (95% CI 0.30–6.11), compatible with substantial benefit or harm and with an upper bound above the overall non-inferiority margin. It does not demonstrate safe omission in intensive care. The more favourable internal jugular estimate and larger subclavian estimate came from small event counts, multiple comparisons and wide intervals. Likewise, the lowest platelet band did not supply a precise estimate of treatment benefit. A higher underlying bleeding risk in a subgroup is not itself evidence that the relative treatment effect differs.

Post-transfusion platelet counts introduce another problem: they were measured after allocation and catheter insertion. Associations between these values and bleeding combine treatment response with patient characteristics and cannot identify a causal preprocedural threshold. The study tested a fixed unit against no prophylactic unit, without verifying an increment before insertion. It did not compare targets of 20, 30 or 50 × 10⁹/L, or one unit against several. Recommendations to increase the dose or achieve a particular count would require additional evidence.²

Finally, the economic result is informative but subordinate to the clinical uncertainty. The reported saving reflects the assigned unit, subsequent blood products and selected bleeding-related costs in this setting. It is not a comprehensive measure of quality-adjusted survival or patient preferences. Generalisability also depends on prompt rescue transfusion, operator expertise and local blood-product costs. Patients with counts below 10 × 10⁹/L, uncorrected major coagulopathy or continued therapeutic anticoagulation were not represented by the principal eligibility range. The trial was publicly funded by ZonMw, but that fact alone neither establishes nor removes other potential biases.^{1,2}

Body of Evidence

Earlier observational work helps explain why PACER was necessary. Zeidler and colleagues retrospectively examined 604 central venous catheter insertions in 193 patients with acute leukaemia undergoing intensive treatment or transplantation. Bleeding was predominantly minor, with no grade 3–4 events; platelet counts below 20 × 10⁹/L were associated with greater bleeding risk. The authors considered insertion without prophylactic transfusion acceptable at counts of at least 20 × 10⁹/L. This provided an empirical basis for restraint, but did not randomise transfusion. Selection, treatment of perceived high-risk patients and retrospective ascertainment limited causal

interpretation. A low observed rate of severe bleeding could not show that prophylaxis had contributed nothing to that rate.³

Cavanna and colleagues reported a particularly reassuring ultrasound-guided internal jugular experience. Among 110 cancer patients with severe thrombocytopenia, 122 procedures were performed without platelet transfusion before or after insertion. There was one self-limiting haematoma and no major bleeding. Nearly all insertions required a single puncture. This demonstrates that uncomplicated access without prophylaxis is possible even at very low counts in a skilled service. It does not establish equivalence to transfusion, since there was no randomised comparator. The exclusively internal jugular setting also differs from PACER's substantial subclavian component. These features are plausible reasons for differences in observed event rates, rather than grounds to discard either study.⁴

Evidence from general prophylaxis in haematological cancer addresses a related but different question. In TOPPS, WHO grade 2–4 bleeding occurred in 50% of patients assigned no prophylaxis and 43% assigned prophylaxis; the no-prophylaxis strategy did not meet its non-inferiority criterion. Platelet use was lower when prophylaxis was withheld. This supports the broader proposition that conserving platelets can trade against bleeding prevention, but the trial concerned prophylaxis during haematological treatment rather than a single catheter insertion. Its event rates, time horizon and clinical circumstances should not be imported into PACER. Equally, a procedural transfusion result should not determine the entire transfusion policy for a subsequent chemotherapy course.⁵

PACER adds the element those procedural series lacked: a contemporaneous randomised comparison of a specific preprocedural intervention. Its prospective grading also captured bleeding requiring prolonged compression, which retrospective series may record inconsistently. The combined evidence therefore supports a more qualified interpretation than either routine transfusion of every thrombocytopenic patient or routine omission whenever ultrasound is available. PACER establishes that, across the enrolled range and mix of procedures, withholding a prophylactic unit increased the broader bleeding outcome. It does not identify a universal numerical threshold or quantify a mortality advantage.¹

For critical care practice at the end of 2023, the result strengthens the case for explicitly considering prophylaxis in a thrombocytopenic patient who requires central venous access. The decision should remain connected to the urgency of access, the planned procedure, the available operator and the consequences of a bleed. These are clinical considerations, not a subgroup rule validated by this trial. If clinicians choose omission in

an individual patient, PACER supplies reason for vigilance and ready access to rescue treatment; its ICU subgroup cannot be cited as proof that this approach is non-inferior. Conversely, the data do not justify delaying essential access until an untested platelet target is reached. The contribution is a better informed balance between bleeding prevention and transfusion exposure, with uncertainty about severe events and the optimal threshold still visible.

Can prophylactic platelet transfusion be routinely omitted before ultrasound-guided CVC insertion at platelet counts of $10\text{--}50 \times 10^9/\text{L}$?

PACER does not support routine omission: grade 2–4 catheter-related bleeding increased and non-inferiority was not established. Major-bleeding effects and subgroup-specific thresholds remain uncertain.

References

1. van Baarle FLF, van de Weerd EK, van der Velden WJFM, Ruiterkamp RA, Tuinman PR, Ypma PF, et al. Platelet Transfusion before CVC Placement in Patients with Thrombocytopenia. *N Engl J Med* 2023;388(21):1956–1965.
2. van Baarle FLF, van de Weerd EK, van der Velden WJFM, Ruiterkamp RA, Tuinman PR, Ypma PF, et al. Supplementary Appendix to: Platelet Transfusion before CVC Placement in Patients with Thrombocytopenia. *N Engl J Med* 2023;388:1956–1965.
3. Zeidler K, Arn K, Senn O, Schanz U, Stussi G. Optimal preprocedural platelet transfusion threshold for central venous catheter insertions in patients with thrombocytopenia. *Transfusion* 2011;51(11):2269–2276.
4. Cavanna L, Citterio C, Di Nunzio C, Orlandi E, Toscani I, Ambroggi M. Central venous catheterization in cancer patients with severe thrombocytopenia: Ultrasound-guide improves safety avoiding prophylactic platelet transfusion. *Mol Clin Oncol* 2020;12(5):435–439.
5. Stanworth SJ, Estcourt LJ, Powter G, Kahan BC, Dyer C, Choo L, et al.; TOPPS Investigators. A No-Prophylaxis Platelet-Transfusion Strategy for Hematologic Cancers. *N Engl J Med* 2013;368(19):1771–1780.

STRATUS

Siegal DM, Belley-Côté EP, Lee SF, Hill S, D'Aragon F, Zarychanski R, et al. Small-Volume Blood Collection Tubes to Reduce Transfusions in Intensive Care: The STRATUS Randomized Clinical Trial. JAMA 2023;330(19):1872–1881.

Introduction

Blood testing is integral to intensive care, but the amount collected often exceeds the quantity required for analysis. Repeated sampling can therefore remove a clinically relevant volume of blood from patients already vulnerable to anaemia. Reducing this avoidable loss is attractive because it acts upstream of transfusion: it may preserve the patient's own red cells without asking clinicians to tolerate a lower haemoglobin concentration. The potential benefit accumulates across repeated draws and prolonged admissions, even when the saving from an individual tube is small.

Conventional vacuum tubes commonly collect several millilitres, while modern analytical equipment requires only a fraction of that volume. Small-volume vacuum tubes use less suction to draw less blood while retaining familiar external dimensions and compatibility with laboratory equipment. They can consequently be introduced through ordinary purchasing and stock replacement. This is different from deliberately underfilling a conventional tube, and does not necessarily require a change in the tests clinicians order. The STRATUS investigators proposed that this relatively simple intervention could reduce red blood cell transfusion without producing more inadequate specimens.¹

The proposition is plausible but not self-proving. Bleeding, inflammatory suppression of erythropoiesis, fluid administration, pre-existing anaemia and transfusion thresholds all influence haemoglobin and transfusion exposure. A reduction in sampling volume may be too small to change the decision to give blood in many patients. Conversely, a modest reduction in units used could matter across an entire health system, particularly if the intervention is inexpensive and does not impair diagnostic care.

STRATUS therefore tested a routine-care policy in a large group of Canadian ICUs. Its stepped-wedge design allowed each unit to change once from standard to small-volume tubes, with the order determined randomly. The resulting question was both clinical and organisational: would introducing these tubes into everyday ICU practice reduce transfusion, and could laboratories continue to process the smaller samples reliably?

Synopsis

STRATUS was a pragmatic, open-label, stepped-wedge cluster randomised trial involving 25 adult medical-surgical ICUs in Canada. Participating units had at least 14 beds, could provide invasive ventilation, used standard-volume tubes and could supply electronic data. Fourteen were in Quebec, nine in Ontario and one each in Manitoba and New Brunswick. Two hospitals contributed two geographically separate ICUs with separate storage facilities. All admitted patients were registered under a waiver of individual consent. The study was supported through academic funding, including the Hamilton Academic Health Sciences Organization; that funder had no role in the conduct, analysis or publication decision.¹

The study began on 5 February 2019 and ended on 21 January 2021. A computer-generated schedule assigned the order of transition, initially with two ICUs switching every six weeks. The schedule was held by an unblinded statistician and concealed from investigators and site staff until two weeks before each transition. Brief education, posters and a video prepared nursing and laboratory staff. Standard tubes were then replaced with small-volume tubes through local purchasing and stock systems. After a six-week baseline period, the planned sequence comprised 13 periods, with a one-week washout around each transition excluded from analysis.^{1,2}

The intervention covered routine blood collection except blood cultures and blood gases. The main report described usual standard volumes of 4–6 mL and small volumes of 1.8–3.5 mL, but the supplementary table shows substantial variation by tube and site, including some tube types whose volume did not change. For example, many sites changed haematology tubes from 4 to 2 mL, whereas several already used 2 mL tubes for that test. This was a locally implemented reduction strategy rather than an identical volume reduction for every specimen. Both periods otherwise followed local sampling practice. Transfusion thresholds, discarded blood from vascular lines, test ordering and vascular catheter use were not standardised by the trial.^{1,2}

The pandemic interrupted implementation. Seven ICUs had not yet switched when transitions were paused from 2 March to 17 August 2020. Units continued using their allocated tube type and data collection continued. The investigators state that, without reviewing or analysing the data, they changed the primary analysis to exclude admissions during this pause. Transitions resumed on 18 August according to the randomised order. The complete cohort, including pause-period admissions, became a prespecified secondary analysis. Thus, the pandemic produced an analysis-population amendment rather than a period with no intervention or outcome data.¹

The primary population comprised patients whose ICU admission lasted at least 48 hours, excluding transition washouts and the pandemic pause. This duration criterion aimed to select patients with sufficient exposure to repeated sampling. Patients already admitted when a unit changed tubes remained assigned to the preceding period. Only the first ICU admission was analysed for patients with multiple admissions. Follow-up continued until hospital discharge, death or 30 days after ICU admission, whichever occurred first.

The primary endpoint was red blood cell units transfused per patient during ICU admission. Secondary outcomes included receiving any red cells, haemoglobin change adjusted for transfusion, ICU and hospital length of stay, mortality and specimen insufficiency. Haemoglobin adjustment subtracted 1 g/dL for each transfused unit to estimate the fall that transfusion might otherwise obscure. Insufficient specimens were assessed through laboratory systems for the two commonly used tube categories serving haemoglobin and creatinine testing. Outcomes and demographic data were extracted electronically and merged in a central database.¹

The investigators anticipated a five-day average admission, transfusion in 40% of patients, two units per patient per ICU admission and an intraclass correlation of 0.01. The planned 18,240 patients across 25 ICUs provided 90% power at a two-sided 5% significance level to detect a 2.5% relative reduction, from 2.00 to 1.95 units. The primary analysis used a negative binomial mixed model adjusted for age and sex, with calendar periods as fixed effects and ICU as a random effect. ICU length of stay was an offset, accounting for differing observation durations. Reported least-squares means were expressed per patient at the median ICU stay. The resulting relative risk concerns transfusion counts in this model, not the separate binary outcome of whether a patient received blood.

Of 58 ICUs assessed, 25 participated and all were analysed. There were 51,037 admissions screened. The primary population excluded 23,320 stays shorter than 48 hours, 6,210 pandemic-pause admissions, 305 washout admissions and one requested data removal, leaving 21,201 patients: 10,261 during small-volume use and 10,940 during standard-volume use. The corresponding complete secondary cohort contained 27,411 patients. Mean age was approximately 63 years. Initial haemoglobin was around 10.5 g/dL in women and 11.1–11.2 g/dL in men. Among patients with diagnostic information, injury was less common during small-volume use, 10.6% versus 18.5%.¹

One to three random stock audits per ICU assessed implementation. Four of 50 audits found a small residual quantity of standard tubes, representing 0.02–2.9% of tubes counted in those audits. The reported 92% adherence therefore describes audits

without this finding, rather than direct confirmation of every blood draw. Transfusion was common: 6,362 primary-population patients received red cells.¹

In the primary model, least-squares mean units were 0.72 with small-volume tubes and 0.79 with standard tubes, RR 0.91 (95% CI 0.79–1.05; P=0.19). The reported absolute reduction was 7.24 units per 100 patients per ICU stay (95% CI –3.28 to 19.44). In the larger secondary population, corresponding means were 0.71 and 0.80, RR 0.88 (95% CI 0.77–1.00; P=0.04), with an estimated reduction of 9.84 units per 100 patients (95% CI 0.24–20.76). These were related analyses of overlapping patients, not independent trials.^{1,2}

In the primary population, 2,975 patients during small-volume use, 29.0%, and 3,387 during standard use, 31.0%, received at least one unit; the adjusted RR was 1.01 (95% CI 0.93–1.09). The median transfusion-adjusted haemoglobin fall was 1.40 versus 1.50 g/dL, with adjusted mean difference 0.10 g/dL (95% CI –0.04 to 0.23). This difference was 0.17 g/dL (95% CI 0.05–0.29) in the secondary population. In primary-population laboratory data, insufficient specimens occurred in 42 of 193,695 small-volume samples and 60 of 195,383 standard-volume samples. Adjusted mortality and length-of-stay comparisons did not demonstrate a clear benefit. The investigators concluded that small-volume tubes may reduce transfusion without compromising specimen sufficiency.^{1,2}

Critique

The principal interpretative issue is the difference between a plausible modest benefit and a confirmatory positive trial. STRATUS did not meet the primary significance criterion. Its primary confidence interval spans a clinically relevant reduction and a small increase in modelled transfusion use. The larger secondary cohort produced a similar point estimate and a narrower interval, with nominal P=0.04. This is supportive evidence, but selecting the significant analysis as the definitive result would reverse the declared hierarchy. Nor are the estimates contradictory simply because their P values fall on different sides of 0.05.¹

The pandemic amendment was consequential because it removed 6,210 otherwise eligible admissions from the primary analysis while preserving them in the secondary analysis. The authors explicitly state that this decision preceded data review or analysis, which argues against treating it as a response to an observed treatment result. Nevertheless, it reduced information and changed the balance of calendar periods contributing to the main estimate. Both analyses deserve visibility. The appropriate conclusion is that a modest reduction is possible, with the primary analysis leaving

uncertainty, rather than that the pandemic created a valid reason to disregard that analysis.

A stepped-wedge design is appropriate for a stock policy that is difficult to allocate independently to neighbouring patients. Concealment of the transition schedule limited advance behaviour changes, and sequential roll-out made implementation feasible. However, later calendar time is increasingly associated with small-volume use. Randomising the order of transition does not eliminate secular changes in case mix, transfusion practice or service pressure. Period adjustment is therefore central, particularly during a pandemic. The absence of a statistically significant treatment interaction with pandemic phase or period cannot prove that all time-varying confounding has been removed.¹

The effective information is determined partly by the 25 clusters and their changing composition, rather than simply by the large patient count. The observed intraclass correlation was 0.054, above the assumed 0.01, although the investigators estimated only a small resulting increase in design effect. More revealingly, an alternative generalised estimating equation analysis retained RR 0.91 but widened the 95% confidence interval to 0.60–1.37, with $P=0.63$. A time-decay correlation model gave results similar to the primary model. These checks support stability of the point estimate under several assumptions while showing that precision is sensitive to the analytical framework.²

Treatment effects also varied between ICUs, with an interaction $P=0.002$ in the primary population. This is credible in a trial where baseline tube volumes, reductions achieved, sampling practices and transfusion thresholds differed. Some tube categories were already small and did not change. A single pooled percentage reduction should therefore not be assumed for every hospital. Stock audits demonstrated successful implementation with little residual standard stock, but did not measure each patient's sampled blood volume or every tube used. The study estimates the effect of introducing a policy; it does not establish a precise relation between millilitres saved and units avoided.

The primary endpoint needs a similarly careful description. The model includes length of stay as an offset and reports transfusion use at a standardised duration. It is not equivalent to the probability that an individual avoids transfusion. The latter outcome was not clearly reduced after adjustment, despite lower unadjusted percentages with small tubes. These apparently different results may coexist if an intervention affects units used among recipients more than the number of recipients, but STRATUS does not

establish that mechanism. Neither the crude percentages nor a derived number needed to treat should replace the reported modelled comparison.¹

Transfusion is objectively recorded, but the decision to give it remained open to clinicians who could recognise the intervention period. Brief education may also have increased awareness of blood conservation beyond the tube itself. Electronic capture reduces the burden and potential selectivity of research measurements; it does not blind treatment decisions. The absence of standardised transfusion thresholds improves relevance to usual practice while adding variability and leaving open a behavioural contribution to the result. This matters most when the estimated effect is small.

Selecting admissions lasting at least 48 hours creates a further limitation. Duration is known after exposure begins and is related to both survival and clinical recovery. Restricting on it can affect comparability if treatment influences early discharge or death, although a material early effect from this intervention is less plausible than for an active physiological therapy. Analyses including admissions of any duration gave no clear transfusion reduction, whether the pandemic pause was excluded or included. Those results do not disprove an accumulating effect during longer stays, but constrain claims about benefit across all ICU admissions.²

Routine data also left important covariates incomplete. Diagnosis was available from 20 ICUs and haemoglobin and creatinine from 24. The injury imbalance is relevant because bleeding and transfusion requirements can differ by admission diagnosis. Adjustment for injury among 15,725 patients from 19 ICUs produced a similar estimate, but cannot account for unmeasured severity or other co-interventions in the entire cohort. The primary adjusted model itself had 30 fewer observations because sex was unavailable. Haemoglobin-change analyses used smaller denominators again. Describing electronic collection as complete follow-up would obscure these distinctions.^{1,2}

The haemoglobin result offers limited mechanistic support. A median difference of roughly 0.1 g/dL is small, and transfusion adjustment assumes a fixed increment of 1 g/dL for every unit, despite variation in blood volume, fluid balance and ongoing losses. Values below zero in the derived adjustment were set to zero, adding another assumption. Analyses restricted to patients who were not transfused avoid that arithmetic but condition on a post-allocation event. They should be treated as exploratory support, not independent confirmation of clinical benefit.

The laboratory finding is reassuring within its scope. Insufficient volume was very uncommon in the specified haemoglobin and creatinine samples, and the primary table does not suggest an increase with small tubes. However, the comparison used specimen

counts and a simple proportions test, rather than fully modelling repeated samples within patients and ICUs. More importantly, the trial did not comprehensively assess turnaround time, haemolysis, add-on requests or tests needing greater volumes. It supports routine feasibility for the measured tests, not an unqualified claim that every aspect of laboratory performance is equivalent.¹

Finally, the lack of an adjusted mortality difference should not be converted into proof of patient-level safety or survival benefit. Raw mortality was higher during small-volume periods, while adjusted estimates were close to one, illustrating the importance of cluster and period structure. The trial's strongest contribution is a feasible blood-conservation policy with a possible small transfusion saving. Purchase costs were similar in participating hospitals, but the study did not establish a comprehensive economic benefit. Local laboratory compatibility and implementation performance remain relevant to adoption.

Body of Evidence

The observational foundation illustrates both the rationale and the uncertainty. Jackson Chornenki and colleagues examined 7,273 patients across four Canadian medical-surgical ICUs. Estimated diagnostic blood loss during admission had a median of 213 mL, and nearly half received red cells. Greater cumulative sampling loss was associated with transfusion. This identifies blood collection as a potentially important target, but cannot show that each additional millilitre caused transfusion: more severely ill patients also undergo more testing and may have other reasons to need blood. A randomised intervention is needed to separate those pathways. STRATUS supplied that comparison at substantially greater scale.³

The STRATUS pilot addressed implementation before definitive effectiveness. In one ICU, 165 patients during a six-week standard-tube period were compared with 204 during a six-week small-tube period. Estimated median daily sampling loss fell from 11 to 6 mL. Audits found complete compliance, inadequate specimens occurred in 0.2% in each period, and focus groups identified no implementation barrier. This was a prospective before-and-after feasibility study, not randomised proof of reduced transfusion. Although the full report appeared in 2023, its results had been presented in 2018 and informed the larger programme. The useful message was that a policy change could achieve a practical reduction in collected volume without obvious disruption.⁴

A separate 2020 pilot randomised trial by Barreda Garcia and colleagues compared paediatric-size and adult tubes in 200 non-bleeding medical ICU patients with haemoglobin of at least 7 g/dL. Transfusion or a fall below 7 g/dL occurred in 6% versus

11%, HR 0.69 (95% CI 0.25–1.90). The estimate was imprecise and the trial was intended to assess feasibility and potential effectiveness. Its paediatric-tube process involved additional handling, whereas STRATUS used reduced-vacuum tubes compatible with established equipment. This distinction matters when judging staff workload and transferring feasibility findings. The earlier trial supports the direction of the biological hypothesis without providing definitive confirmation of transfusion benefit.⁵

Together, these studies distinguish three questions: whether diagnostic sampling removes substantial blood, whether smaller collection volumes can be introduced successfully, and whether that change improves clinical outcomes. The first two have stronger support than the third. STRATUS moves beyond a small feasibility study by randomising an ICU policy across many sites, but its primary confidence interval and model sensitivity retain uncertainty about the size of the transfusion effect. The supplementary findings should refine that judgement rather than substitute for it.^{1,2}

The 2023 evidence therefore supports considering small-volume tubes as one component of blood conservation, especially where current tube volumes leave substantial scope for reduction. This is a clinical implementation judgement based on the combination of plausibility, feasibility and possible benefit; it is not a claim that the primary trial proved superiority. Tube changes address the volume collected for an ordered test. They do not establish whether the test was needed, prevent all blood discarded when sampling from lines, or remove other causes of ICU anaemia.

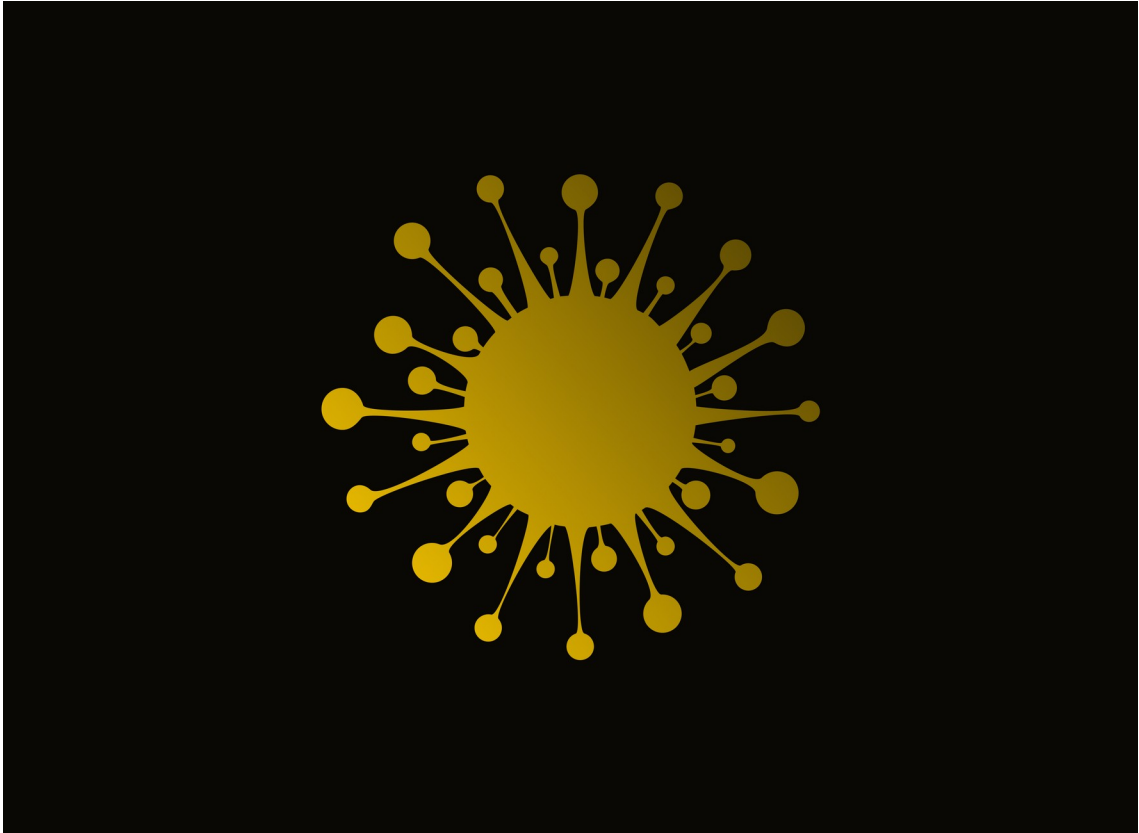
For an ICU considering adoption, the practical implication is to assess the actual local volume change and work with the laboratory on the tests affected. The relevant outcomes extend beyond stock replacement to specimen adequacy and clinical workflow. Any expected reduction in transfusion should be described as modest and uncertain, with no demonstrated improvement in survival. At the book's 2023 cut-off, STRATUS offers a credible, implementable means of reducing avoidable blood collection, while leaving the magnitude and distribution of patient benefit incompletely resolved.

Do small-volume blood collection tubes reduce transfusion in adult ICUs?

They may produce a modest reduction, but STRATUS did not demonstrate a significant benefit in its primary analysis. Specimen insufficiency remained rare for the laboratory tests assessed.

References

1. Siegal DM, Belley-Côté EP, Lee SF, Hill S, D'Aragon F, Zarychanski R, et al. Small-Volume Blood Collection Tubes to Reduce Transfusions in Intensive Care: The STRATUS Randomized Clinical Trial. *JAMA* 2023;330(19):1872–1881.
2. Siegal DM, Belley-Côté EP, Lee SF, Hill S, D'Aragon F, Zarychanski R, et al. Supplemental Online Content to: Small-Volume Blood Collection Tubes to Reduce Transfusions in Intensive Care. *JAMA* 2023;330:1872–1881.
3. Jackson Chornenki NL, James TE, Barty R, Liu Y, Rochweg B, Heddle NM, et al. Blood loss from laboratory testing, anemia, and red blood cell transfusion in the intensive care unit: a retrospective study. *Transfusion* 2020;60(2):256–261.
4. Siegal DM, Belley-Côté EP, Lee SF, Robertson T, Hill S, Benoit P, et al. Small-volume tubes to reduce anemia and transfusion (STRATUS): a pilot study. *Can J Anaesth* 2023;70:1797–1806.
5. Barreda Garcia J, Xian JZ, Pedroza C, Salahuddin M, Mak G, Keene A, et al. Pediatric size phlebotomy tubes and transfusions in adult critically ill patients: a pilot randomized controlled trial. *Pilot Feasibility Stud* 2020;6:112.



Sepsis Trials

REMAP-CAP Simvastatin

The REMAP-CAP Investigators. Simvastatin in Critically Ill Patients with Covid-19. N Engl J Med 2023;389(25):2341–2354.

Introduction

An inexpensive, widely available drug that reduced the consequences of severe viral pneumonia would have considerable value. Simvastatin was a plausible candidate because its biological effects extend beyond lowering cholesterol. Experimental work suggested effects on inflammation and lung injury, while observational associations between antecedent statin use and outcomes encouraged interest in repurposing the drug for coronavirus disease 2019 (COVID-19). Those associations could not establish whether starting treatment during critical illness would help. Patients already receiving statins differ from untreated patients in comorbidity, access to care and other medication use.¹

The question also evolved during the pandemic. By the time much of this trial was undertaken, glucocorticoids and interleukin-6 receptor antagonists had become common background treatments in the participating ICUs. Simvastatin therefore needed to provide an additional benefit in patients receiving other anti-inflammatory therapies. A mechanism that appeared promising in experimental lung injury did not guarantee an appreciable effect in that setting.

The Randomized, Embedded, Multifactorial Adaptive Platform Trial for Community-Acquired Pneumonia, REMAP-CAP, offered a means of testing this proposition while simultaneously evaluating other interventions. Its pandemic stratum included patients with suspected or confirmed COVID-19 and distinguished critically ill patients receiving organ support from those with less severe illness. Adaptive allocation could direct more patients towards a promising intervention as information accumulated, while a common statistical framework accounted for other domain assignments.¹

The simvastatin comparison is instructive because its result lies between convincing confirmation and confident exclusion of benefit. Recruitment ended as COVID-19 cases and enrolment declined, before the planned statistical criteria had been reached. Several estimates favoured treatment, but the trial's success threshold remained unmet. Appraisal therefore requires attention to the size and uncertainty of the effect, the changing composition of the comparison groups and the accompanying adverse events.

The clinical question is specifically whether to initiate high-dose simvastatin for severe COVID-19 in patients who were not already receiving a statin.

Synopsis

This was an international, open-label, randomised comparison within the ongoing REMAP-CAP platform. The simvastatin domain recruited at 141 sites in 13 countries between 28 October 2020 and 8 January 2023. Adults admitted with suspected or confirmed COVID-19 were eligible. Critical illness required ICU treatment with respiratory or cardiovascular organ support. Qualifying respiratory support included invasive or non-invasive ventilation, or high-flow nasal oxygen at at least 30 L/min with an inspired oxygen fraction of at least 0.4. Cardiovascular support comprised vasopressor or inotrope infusion. The platform obtained local ethical approval and consent through locally approved arrangements. Public and charitable organisations supported the study; funders had no role in its design, analysis or publication decision.^{1,2}

Important exclusions included more than 48 hours since ICU admission or qualifying sustained organ support, severe liver disease, simvastatin hypersensitivity, incompatible medication and creatinine above 200 µmol/L in patients not receiving renal replacement therapy. Patients currently receiving a statin, or for whom the clinician intended to start one, were excluded from this domain. Platform exclusions also addressed imminent death with limits to active treatment, imminent discharge and prolonged hospital admission. These restrictions selected a population suitable for new statin treatment rather than an unselected COVID-19 ICU cohort.²

Patients were allocated through a central computerised system to enteral simvastatin 80 mg once daily or no statin. The first dose was intended as soon as possible, ideally within four hours. Treatment continued until first ICU discharge or day 28, whichever came first. Other care followed clinical practice and any additional randomised platform assignments. Liver enzymes, creatine kinase and renal function were to be checked during days 7–14 and again during days 21–28 if treatment continued. There was no placebo; patients, treating clinicians and outcome assessors knew the allocation. Aggregate results and adaptive allocation probabilities remained concealed from the investigators responsible for clinical recruitment.^{1,2}

Allocation began equally. Following the first adaptive update in April 2021, critically ill patients were assigned in a 9:1 ratio favouring simvastatin. This continued through subsequent updates before the control allocation minimum increased to 25% in September 2022. Thus, the final group sizes and their distribution across time were intentionally unequal. The trial had no fixed maximum sample size. Superiority required a

posterior probability above 99% that the treatment odds ratio exceeded 1; statistical futility required a probability above 95% that it was below 1.2. Recruitment stopped for operational reasons, following simulations suggesting that declining enrolment made a statistical conclusion unlikely within a feasible period. The steering committee made that decision before unblinding.^{1,2}

The primary endpoint was organ support-free days through day 21, an ordinal composite incorporating survival. In-hospital death before discharge and within 90 days was assigned -1, including deaths after day 21. Survivors were ranked according to days free of respiratory and cardiovascular support. In critically ill patients the scale ran from -1 to 21. This endpoint therefore gives death the worst rank and uses support duration to distinguish other outcomes. It is not simply the number of ventilator-free days among survivors.

A Bayesian cumulative logistic model adjusted for age, sex, site nested within country, calendar epochs, domain eligibility and other randomised assignments. Effects were estimated separately for critically ill and noncritically ill patients, with hierarchical borrowing between severity groups. The overall log-odds treatment prior was centred on no effect and weakly informative. Other platform participants contributed to estimation of covariate effects, but only patients allocated to no statin within this domain served as controls for simvastatin efficacy. The model contained no simvastatin interaction terms with other treatments in its primary specification. Higher odds ratios denoted a more favourable outcome.^{1,2}

Of 2,739 critically ill patients randomised, 1,884 were assigned simvastatin and 855 control. Consent withdrawal removed 38 and 13 respectively. Four further patients lacked primary outcomes, leaving 1,843 and 841 in the primary comparison. The baseline and safety population comprised 2,688 patients. A further 184 noncritically ill participants contributed domain data to the hierarchical model. Follow-up concluded on 15 April 2023. Missing outcomes were excluded from the relevant analyses rather than imputed; unknown follow-up status in time-to-event analyses could be censored.^{1,2}

Median age was 56 versus 57 years, and approximately two-thirds were male. Around one-third were receiving invasive ventilation; the remainder predominantly received non-invasive ventilation or high-flow oxygen. Median ICU-to-enrolment time was approximately 17 hours. Europe contributed 81.6% of simvastatin recipients and 92.8% of controls. Glucocorticoids were given to 97.2% and an interleukin-6 receptor antagonist to 52.4% overall. At least one simvastatin dose was recorded in 1,744 of 1,846 assigned patients, 94.5%; 832 of 842 controls, 98.8%, met the trial's no-treatment

adherence definition. Among patients with treatment-duration data, median exposure within the intended treatment window was seven days.^{1,2}

Median organ support-free days were 11 (IQR –1 to 17) with simvastatin and 7 (IQR –1 to 16) with control. The adjusted odds ratio was 1.15 (95% credible interval 0.98–1.34), giving a 95.9% posterior probability of superiority. This did not meet the 99% success threshold. Survival to hospital discharge was 1,352/1,843, 73.4%, versus 589/841, 70.0%; the adjusted survival odds ratio was 1.04 (95% credible interval 0.85–1.27), with a 64.4% probability of benefit. The probability that the primary odds ratio was below 1.2 was 71.8%, also short of the statistical futility criterion.^{1,2}

At 90 days, death had occurred in 504/1,835 patients with known status receiving simvastatin, 27.5%, and 257/837 controls, 30.7%. The reported hazard ratio, parameterised in the favourable survival direction, was 1.12 (95% credible interval 0.95–1.32), with 91.9% probability of benefit. Among patients not receiving invasive ventilation or extracorporeal membrane oxygenation at baseline, progression to either support or death occurred in 37.0% versus 42.5%, favourable odds ratio 1.23 (0.98–1.55). Respiratory support-free days through day 28 favoured simvastatin, OR 1.16 (1.00–1.35), while the modified WHO status scale at day 14 gave OR 1.23 (1.06–1.43).¹

Serious adverse events were reported in 57/1,846 simvastatin patients, 3.1%, and 17/842 controls, 2.0%. The reported adjusted odds ratio for an adverse event was 1.56 (95% credible interval 1.13–2.14). Thirteen simvastatin patients had elevated aminotransferases and thirteen elevated creatine kinase; none were reported in controls. Many of these events prompted treatment cessation. The investigators concluded that simvastatin had not met the prespecified criteria for superiority, despite estimates suggesting possible benefit.^{1,2}

Critique

The central issue is how to interpret a high probability of benefit when the trial's confirmatory threshold was not reached. A posterior probability of 95.9% means that the fitted model assigns substantial probability to an odds ratio above 1. It is neither a 95.9% chance that each treated patient benefits nor proof of an effect large enough to justify routine treatment. The credible interval includes no effect and ranges from a small adverse shift to a potentially worthwhile favourable shift. Respecting the 99% threshold prevents a favourable but incomplete result from being promoted into a successful trial. Equally, missing that threshold does not demonstrate that simvastatin is ineffective.¹

Operational closure must be distinguished from statistical futility. The domain ended because recruitment had become too slow to make a timely statistical conclusion likely, rather than because the posterior distribution met the futility boundary. The final primary futility probability was only 71.8%. The decision was reasonable in the circumstances, but left residual uncertainty. There is no evidential basis for assuming that additional recruitment would necessarily have confirmed benefit; later observations could also have moved the estimate towards no effect or harm.²

Response-adaptive allocation adds a specific limitation. Preferentially assigning treatment to the apparently better arm may benefit participants, but reduces information from the smaller comparator group. More than four-fifths of controls entered during the initial allocation period, compared with little over one-third of simvastatin patients. Consequently, crude comparisons mix treatment with changes in viral epidemiology, clinical care, recruitment geography and ICU pressure. Adjustment for time and site is essential, and the apparent reduction in unadjusted mortality cannot be interpreted independently of the adjusted model. Randomisation remains valuable, but unequal allocation across a changing pandemic does not create the same simple comparison as constant 1:1 randomisation.²

The platform analysis also needs an accurate description. Thousands of other participants were included to estimate prognostic and treatment covariates; they were not an enlarged untreated simvastatin control group. Within-domain concurrent controls anchored the treatment comparison. Borrowing across severity states was an additional modelling choice, but the noncritical domain sample was small. Sensitivity analyses without that borrowing produced similar critical illness estimates. Restricting the model to the simvastatin domain gave OR 1.14 (0.98–1.35), close to the main estimate, whereas using other closed domains gave OR 1.18 (1.01–1.37). These analyses suggest a reasonably stable direction, while showing that whether the interval excludes 1 depends partly on specification. None should replace the primary model because it gives a more attractive result.²

The completed primary analysis report describes imperfect convergence across some simulation chains, attributed to confounding among site, country and time parameters. It also states that the treatment-effect distributions were consistent across chains. This is a qualified reassurance: instability in nuisance parameters is less concerning than instability in the treatment contrast, but it reinforces the importance of the sensitivity analyses and limits a claim that a complex model is automatically more reliable. The simpler within-domain analysis is particularly useful here because its estimate remains close to the primary result.²

The ordinal endpoint efficiently combines death and duration of support, but its interpretation is easily overstated. A four-day difference between group medians does not mean simvastatin gives each patient four additional days free of support. The adjusted common odds ratio summarises shifts across multiple cut-points, including survival. The survival component itself had an adjusted odds ratio close to 1. Dichotomised sensitivity analyses varied across the support-free-day scale, with generally more favourable estimates at some higher cut-points. The overall result therefore cannot be reduced to an established mortality benefit or a uniform acceleration of recovery.^{1,2}

Open-label treatment creates another route to bias. Starting, escalating and withdrawing respiratory support involve judgement, especially for non-invasive support, and clinicians knew whether simvastatin had been prescribed. Death is less susceptible to ascertainment bias, but support duration and discharge are more vulnerable to management effects. The survival and resource-use components should consequently be considered together. The near-complete primary outcome dataset is reassuring, although consent withdrawals occurred after allocation and cannot simply be counted as successful follow-up.

Several secondary outcomes point in the same favourable direction, including respiratory support-free days, progression to invasive support or death, and day-14 clinical status. However, they are correlated measures in largely the same patients, not independent replications. The WHO result had a posterior probability above 99%, but success on a secondary scale cannot retrospectively satisfy the primary criterion. Length-of-stay outcomes were also defined to rank deaths unfavourably, assigning a 90-day duration to deaths in the relevant setting. Their discharge estimates should not be read as uncomplicated differences in bed-days among survivors.²

The safety signal matters because the expected benefit is uncertain. Familiarity with statins in ambulatory cardiovascular practice does not establish the safety of initiating 80 mg simvastatin during critical illness. Open-label allocation and targeted monitoring may have influenced detection and attribution of adverse events, but do not justify dismissing the excess. The supplement distinguishes 62 events from 57 affected simvastatin patients and documents creatine kinase elevations accompanied by acute kidney injury in several cases. A sensitivity analysis adding calendar-time adjustment reduced the estimated probability of harm from 99.6% to 89.3% and widened the interval across no effect. This qualifies certainty about the magnitude of harm without removing the observed events. The practical burden includes laboratory surveillance, interaction checks and treatment cessation when toxicity develops.²

Adherence was adequate for estimating a policy of initiation, but its definition was permissive: a single dose qualified as adherence in the active group. It does not establish delivery of the intended daily course. Stopping at ICU discharge appropriately limits exposure, while clinician decisions, intolerance and incomplete enteral administration may further alter dose received. A per-protocol analysis that selects patients who tolerated or received treatment cannot restore the balance created by randomisation and should not displace the intention-to-treat result.

Finally, the subgroup findings do not establish a prescribing rule. Signals among patients not invasively ventilated, or with higher inflammatory markers, arose from numerous comparisons. Ferritin was missing in many participants. The intended inflammatory phenotype analysis could not be meaningfully performed because 98.8% were classified as hypoinflammatory. This limits translation of a phenotype hypothesis from other forms of acute respiratory distress syndrome. The study also excludes established statin users and many patients with renal, hepatic or medication contraindications, and predominantly reflects high-income European practice. It cannot answer whether to continue an indicated chronic statin, or establish effectiveness in a substantially different later pandemic population.²

Body of Evidence

The most direct pre-pandemic comparison was HARP-2, which tested enteral simvastatin 80 mg daily against placebo in 540 patients enrolled within 48 hours of acute respiratory distress syndrome onset. Its primary endpoint was ventilator-free days through day 28. Means were 12.6 versus 11.5 days, with $P=0.21$; mortality at day 28 was 22.0% versus 26.8%, with $P=0.23$. This trial did not establish benefit in an unselected ARDS population despite numerically favourable estimates. Its placebo control and similar dose make it relevant, but the underlying diseases and outcome definition differ from the COVID-19 platform comparison.³

A subsequent secondary analysis of 539 HARP-2 participants identified hypoinflammatory and hyperinflammatory subphenotypes. The latter group had substantially worse outcomes and better observed survival with simvastatin than placebo. This generated a credible hypothesis that biological heterogeneity could conceal treatment responsiveness within an overall inconclusive trial. It did not constitute a new randomised trial in prospectively selected patients. A favourable comparison within a subgroup also requires care when distinguishing prognostic classification from a reproducible treatment interaction. The appropriate next step was prospective testing of a selection strategy, rather than assuming that any raised inflammatory marker identified a simvastatin responder.⁴

SAILS provided a less encouraging class comparison. In 745 patients with sepsis-associated ARDS, rosuvastatin did not reduce mortality before discharge home or day 60, 28.5% versus 24.9%, and ventilator-free days were essentially identical. Recruitment stopped for futility. Fewer days free of hepatic and renal failure raised concern about organ dysfunction. Rosuvastatin and simvastatin are different drugs, so this is not a direct refutation of the REMAP-CAP result. It does show why anti-inflammatory plausibility and extensive cardiovascular experience cannot substitute for critical illness efficacy and safety data.⁵

The COVID-19-specific INSPIRATION-S trial tested atorvastatin 20 mg daily against placebo in 11 Iranian hospitals. Of 605 randomised patients, 587 entered the primary analysis. The 30-day composite of venous or arterial thrombosis, extracorporeal membrane oxygenation or death occurred in 33% versus 36%, OR 0.84 (95% CI 0.58–1.21). Death occurred in 31% versus 35%, again with substantial uncertainty. This double-blind study did not establish efficacy, although its interval allowed a clinically important benefit. It used a different statin, lower-dose strategy and different composite endpoint, and therefore cannot be treated as a smaller duplicate of REMAP-CAP.⁶

Together, these trials support a narrower interpretation than either a beneficial statin class effect or definitive class failure. HARP-2 and REMAP-CAP leave room for a modest simvastatin effect, while SAILS demonstrates that results and adverse effects may differ across compounds and populations. INSPIRATION-S adds a direct COVID-19 comparison without providing confirmatory benefit. Pooling unlike regimens may help assess broad consistency but cannot identify which dose, disease stage or biological subgroup should be treated. None of these studies establishes that antecedent statin associations can be reproduced by starting a statin after ICU admission.^{1,3,5,6}

REMAP-CAP makes a substantial contribution through its larger COVID-19 sample and extensive use of concomitant glucocorticoids and interleukin-6 blockade. It addresses whether simvastatin adds to a more developed treatment background. Its results preserve a research hypothesis, particularly concerning respiratory recovery, but the unmet primary threshold, model dependence and observed toxicity prevent a straightforward recommendation for routine initiation. Evidence available by the end of 2023 supports further targeted evaluation, with a prospectively defined population and careful safety monitoring. Established cardiovascular indications and continuation of chronic therapy remain separate decisions, because this comparison deliberately excluded those patients.^{1,2}

Should simvastatin be started routinely in critically ill patients with COVID-19?

Routine initiation is not supported. Possible benefit did not meet the prespecified superiority threshold, and serious adverse events were more frequent.

References

1. The REMAP-CAP Investigators. Simvastatin in Critically Ill Patients with Covid-19. *N Engl J Med* 2023;389(25):2341–2354.
2. The REMAP-CAP Investigators. Supplementary Appendix to: Simvastatin in Critically Ill Patients with Covid-19. *N Engl J Med* 2023;389:2341–2354. Updated 21 December 2023.
3. McAuley DF, Laffey JG, O’Kane CM, Perkins GD, Mullan B, Trinder TJ, et al. Simvastatin in the Acute Respiratory Distress Syndrome. *N Engl J Med* 2014;371:1695–1703.
4. Calfee CS, Delucchi KL, Sinha P, Matthay MA, Hackett J, Shankar-Hari M, et al. Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial. *Lancet Respir Med* 2018;6(9):691–698.
5. National Heart, Lung, and Blood Institute ARDS Clinical Trials Network. Rosuvastatin for Sepsis-Associated Acute Respiratory Distress Syndrome. *N Engl J Med* 2014;370(23):2191–2200.
6. INSPIRATION-S Investigators. Atorvastatin versus placebo in patients with covid-19 in intensive care: randomized controlled trial. *BMJ* 2022;376:e068407.

LOVIT-COVID

The LOVIT-COVID Investigators, on behalf of the Canadian Critical Care Trials Group, and the REMAP-CAP Investigators. Intravenous Vitamin C for Patients Hospitalized With COVID-19: Two Harmonized Randomized Clinical Trials. JAMA 2023;330(18):1745–1759.

Introduction

COVID-19 created an urgent need for treatments that could prevent respiratory deterioration and improve survival across different healthcare systems. Vitamin C was an attractive candidate because it was widely available and had a plausible role in modifying oxidative stress and vascular injury. Low circulating concentrations in severe illness added to the rationale. However, a low concentration does not establish that pharmacological replacement reverses the processes causing organ failure, and widespread availability cannot substitute for evidence of clinical benefit.¹

The preceding critical care literature offered conflicting reasons for enthusiasm and caution. CITRIS-ALI had not improved its primary organ dysfunction or biomarker outcomes, although a secondary mortality signal supported further evaluation. Trials of combinations containing vitamin C addressed a different treatment strategy. When the COVID-19 studies were designed, the consequences of administering high-dose vitamin C alone to patients with predominantly viral respiratory failure remained uncertain.²

That uncertainty changed during recruitment. The 2022 LOVIT trial found more death or persistent organ dysfunction with intravenous vitamin C among patients with sepsis receiving vasopressors. This was relevant external evidence, but did not directly answer the question for hospitalised COVID-19 patients with respiratory failure or for those who had not yet required critical care. It also made safety monitoring and the criteria for continuing recruitment especially important.³

The investigators of LOVIT-COVID and the vitamin C domain of REMAP-CAP agreed before recruitment to harmonise their treatment and outcomes and ultimately combine their data. The resulting comparison examined the same four-day intravenous regimen in critically ill and noncritically ill patients. The hypothesis was that vitamin C would increase survival and time free of organ support; the challenge was to assess this efficiently while retaining the differences between two independently managed trials.^{1,4}

Synopsis

The report combined a Canadian blinded, placebo-controlled trial with an international, open-label adaptive platform domain. Recruitment ran from 23 July 2020 to 15 July 2022, with critically ill participants recruited at 90 sites and noncritically ill participants at 40 sites on four continents. These site counts overlap. LOVIT-COVID was supported by the Lotte and John Hecht Memorial Foundation, which also supported REMAP-CAP; the platform had additional public and charitable funding. The funders had no role in study conduct, analysis or publication decisions. Separate steering committees and safety boards oversaw the trials, with information exchanged between boards.¹

Eligible participants were hospitalised adults with COVID-19, confirmed in LOVIT-COVID and suspected or proven in REMAP-CAP. Critical illness required ICU admission and respiratory or cardiovascular support. Respiratory support included invasive or noninvasive ventilation or high-flow nasal oxygen at at least 30 L/min and an inspired oxygen fraction of at least 0.4; cardiovascular support meant a vasopressor or inotrope infusion. Patients receiving support outside the ICU could therefore fall within the noncritical stratum. LOVIT-COVID excluded patients who had received vasopressors during the admission. Both trials excluded hospitalisation exceeding 14 days for the acute illness. LOVIT-COVID also excluded more than 24 hours of noninvasive or invasive respiratory support; REMAP-CAP excluded more than 24 hours since ICU admission. Other important exclusions included imminent death, known glucose-6-phosphate dehydrogenase deficiency, vitamin C allergy, recent symptomatic kidney stones and previous intravenous vitamin C other than nutritional supplementation.^{1,4}

Allocation used concealed computer-generated schedules. LOVIT-COVID assigned vitamin C or placebo in a fixed 1:1 ratio, stratified by site. REMAP-CAP stratified allocation by illness state, initially equally, then adjusted probabilities according to accumulating results. Its control was usual care without vitamin C. Informed consent was obtained from patients or surrogates, before or after allocation according to local requirements. Vitamin C was given intravenously at 50 mg/kg over 30–60 minutes every six hours for 96 hours, to a maximum of 16 doses. Products were procured locally. An unblinded pharmacist prepared study treatment in LOVIT-COVID; its patients and treating teams remained blinded, whereas treatment was open-label in REMAP-CAP. Other care remained at clinicians' discretion, and glucose monitoring precautions addressed interference with some bedside assays.^{1,4}

The primary endpoint combined survival to hospital discharge with days alive and free of respiratory and cardiovascular support in the ICU through day 21. In-hospital death, with follow-up censored at 90 days, received the worst score of –1. Survivors were ranked by time free of support; noncritically ill patients who never required support received the

best score of 22. Thus the endpoint was an ordinal clinical ranking rather than an ordinary duration. Secondary outcomes included survival without persistent organ dysfunction at day 28, 90-day survival, respiratory and vasopressor-free days, intubation and live discharge. Persistent dysfunction meant invasive ventilation, vasopressor treatment or new kidney replacement therapy.^{1,4}

LOVIT-COVID had originally planned 800 participants to detect a nine-percentage-point reduction from a 30% risk of death or persistent organ dysfunction, with 80% power. The harmonised comparison adopted the platform endpoint and enrolment strategy; REMAP-CAP had no fixed maximum sample. Bayesian cumulative logistic models estimated separate treatment effects by severity, with a neutral overall prior and hierarchical borrowing between strata. Adjustment included trial, age, sex, site within country, calendar periods and eligibility and allocation within platform domains. Patients outside the vitamin C domain informed covariate effects, not the vitamin C treatment comparison. Odds ratios above 1 favoured vitamin C.^{1,4}

Original decision thresholds were posterior probabilities above 99% for efficacy or inferiority, and above 90% for equivalence within odds ratios of 1/1.2 to 1.2. Following the LOVIT results, thresholds were added for futility, above 95% probability of an odds ratio below 1.2, and harm, above 90% probability of an odds ratio below 1. The two trials stopped after REMAP-CAP met harm and futility criteria in both strata. Their interim analyses remained separate; the pooled analysis followed completion of both trials.¹

Of 2,613 randomised participants, seven were classified as ineligible, 15 withdrew consent for follow-up and one critical control participant lacked the primary outcome. The primary analysis included 2,590 patients: 2,206 from REMAP-CAP and 384 from LOVIT-COVID. There were 1,568 critically ill patients, assigned vitamin C in 1,037 cases and control in 531, and 1,022 noncritically ill patients, assigned vitamin C in 456 cases and control in 566. Two included patients allowed use of previously collected data after withdrawing follow-up consent; their last known status was carried forward for the primary endpoint. Other missing primary outcomes were not imputed.¹

Median age was 60 years in the critical stratum and 62 years in the noncritical stratum. Among critically ill participants, 28.0% were invasively ventilated, 36.2% received noninvasive ventilation and 35.1% received high-flow oxygen. Most noncritically ill patients received low-flow oxygen or no respiratory support. Corticosteroids were used in 95.1% overall. In LOVIT-COVID, 96.1% received at least 90% of scheduled doses. In REMAP-CAP, 95.2% of vitamin C recipients had no treatment-delivery deviation; 13 received none of the assigned treatment, and six controls received intravenous vitamin C.^{1,4}

In critically ill patients, median organ support-free days were 7 (interquartile range –1 to 17) with vitamin C and 10 (–1 to 17) with control. The adjusted proportional odds ratio was 0.88 (95% credible interval 0.73–1.06), with posterior probabilities of 8.6% for efficacy, 91.4% for harm and 99.9% for futility. In noncritically ill patients, medians were 22 (18–22) and 22 (21–22), respectively. The adjusted odds ratio was 0.80 (0.60–1.01), with probabilities of 2.9%, 97.1% and greater than 99.9%, respectively.¹

Hospital survival was 642/1,037 (61.9%) versus 343/531 (64.6%) in the critical stratum, adjusted odds ratio 0.92 (0.73–1.17). In the noncritical stratum it was 388/456 (85.1%) versus 490/566 (86.6%), odds ratio 0.86 (0.61–1.17). The adjusted survival differences were –1.9 percentage points in each stratum, with credible intervals of –7.5 to 3.5 and –6.8 to 1.7, respectively. Ninety-day survival was 59.8% versus 62.1% and 81.5% versus 82.8%; the reported survival hazard ratios were 0.94 (0.80–1.11) and 0.93 (0.74–1.19). Survival without persistent organ dysfunction at day 28 also did not favour vitamin C.¹

Intubation occurred in 266/750 (35.5%) versus 124/381 (32.5%) critically ill patients not already intubated, and in 63/456 (13.8%) versus 50/566 (8.8%) noncritically ill patients. Other secondary estimates offered no convincing benefit. Serious adverse events affected 27/1,493 (1.8%) vitamin C recipients and 9/1,098 (0.8%) controls. Four events were considered possibly or probably vitamin C-related: methemoglobinaemia, two episodes of hypoglycaemia and haemolytic anaemia in a patient subsequently found to have glucose-6-phosphate dehydrogenase deficiency. Sensitivity analyses broadly supported the primary findings; no persuasive subgroup with benefit emerged.^{1,4}

Critique

The clinically decisive finding is the extremely low probability of a worthwhile improvement with this regimen, accompanied by a concerning direction of effect. Futility here meant an odds ratio below 1.2, not proof of an exactly zero effect. The posterior probabilities of harm exceeded 90%, but the 95% credible intervals for the primary estimates still included 1. These statements are compatible because they express different probability thresholds. The trial argues strongly against routine treatment without proving a precise increase in mortality or excluding every very small benefit.¹

The two severity strata strengthen the clinical scope, but should not be collapsed into a single crude comparison. Their allocation ratios, baseline risks and endpoint distributions differ considerably. Death and duration of support both influence the critical-stratum result; in the noncritical stratum, many patients occupy the maximum score. Equal medians of 22 therefore conceal differences in deterioration and death. Conversely, the critical-stratum median separation of three days is not an estimated three-day causal

prolongation of support for every treated patient. Hospital survival and the ordinal effect should be interpreted together.^{1,4}

Harmonisation increased the amount of directly relevant randomised evidence and used the same dose and common outcome definitions. It was agreed before recruitment, rather than being an opportunistic combination of completed positive studies. Nevertheless, the resulting report is dominated by the 2,206 REMAP-CAP participants. The 384-patient placebo-controlled component cannot provide a comparably precise independent check. LOVIT-COVID also excluded vasopressor exposure and recruited older patients than the corresponding platform groups. Adjustment for trial accounts for measured differences in overall prognosis; it does not make recruitment, care or measurement identical.^{1,4}

This matters when judging agreement between trials. Supplementary analyses of LOVIT-COVID alone were imprecise, and its hospital-survival point estimates favoured vitamin C, although with wide credible intervals crossing 1. The between-trial coefficients in the supplementary overall-outcome table describe prognosis in one trial relative to the other, not a definitive test that treatment effects are identical. The defensible conclusion is that the smaller trial does not establish a contradictory benefit. It is too strong to claim that two equally persuasive independent trials each demonstrated harm.⁴

Response-adaptive randomisation introduces a particularly important operational qualification. Early REMAP-CAP analyses favoured vitamin C in critically ill patients and increased its allocation probability. A ten-month interval followed the penultimate adaptive analysis while new international data-flow processes were implemented; more than half of recruitment occurred during this interval. Eventually 69% of critically ill platform participants had been assigned vitamin C. The design remained randomised, but adapting allocation to unstable early estimates did not ensure that more participants received the ultimately preferable treatment.^{1,4}

Calendar adjustment was therefore essential. Recruitment geography, clinical practice and viral epidemiology changed while allocation probabilities differed. The primary model used short time periods, location and other covariates, and sensitivity analyses restricted the populations informing adjustment. Results remained broadly unfavourable. Exploratory effects nevertheless varied markedly across allocation periods, with an early favourable period followed by unfavourable periods. Country-level variant assignments and continent analyses did not resolve that variation. These observations justify caution about the exact effect magnitude; selecting the most favourable interval would abandon the overall randomised question.^{1,4}

The Bayesian model also requires careful explanation. Its 9,802 primary outcomes do not represent 9,802 vitamin C randomisations: most additional patients inform covariate estimation through other platform domains. The direct comparison remains between eligible concurrently randomised vitamin C and control participants. Hierarchical borrowing can improve precision, but makes the severity-specific estimates partly dependent on information in the other stratum. Sensitivity analyses without borrowing and analyses confined to the vitamin C comparison remained unfavourable, which supports the central conclusion without making model assumptions irrelevant.^{1,4}

The changes to stopping thresholds were prompted by external evidence of harm. This is a clinically understandable response, but the final harm and futility criteria should not be presented as the original decision rules. Neither primary estimate met the original 99% inferiority threshold. The changed criteria permitted stopping when a useful benefit had become highly unlikely and possible harm was credible. The appropriate appraisal rests on the complete probability distributions and supporting outcomes, with that history explicit, rather than on a binary label of a positive or negative trial.¹

Open-label treatment could influence decisions to intubate, continue respiratory support or discharge a patient. This concern is more relevant to the ordinal endpoint and intubation findings than to the occurrence of death. It does not explain the findings automatically, especially as survival estimates also pointed away from benefit. Similarly, the many secondary comparisons should be treated as supporting evidence rather than independent definitive demonstrations of specific harms. The high adherence and very limited control crossover make dilution through failure to deliver vitamin C an unconvincing explanation for the absence of benefit.^{1,4}

Follow-up was strong for the primary endpoint, but “no imputation” needs qualification. Two primary statuses used last observation carried forward. For later survival outcomes, the supplementary definitions also permitted assuming survival after discharge alive without support when subsequent mortality status had not been recorded. Reported crude 90-day percentages excluded patients censored before that point, whereas time-to-event models used their available follow-up. These conventions do not appear sufficient to rescue efficacy, but preclude claiming independently observed 90-day status for every participant.^{1,4}

Safety ascertainment was not identical across trials. All reported serious adverse events arose in REMAP-CAP, while LOVIT-COVID specifically collected hypoglycaemia and screened for haemolysis. Many platform events were related to another assigned intervention or general trial participation; the overall count is not a count of proven vitamin C toxicity. Nevertheless, the possibly related events show that a familiar vitamin

can cause clinically consequential problems at pharmacological doses. The mechanism behind the overall unfavourable efficacy estimate remains unknown, and the limited safety data cannot identify it.⁴

Applicability is strongest to early treatment of hospitalised COVID-19 in the corticosteroid era, across a broad range of respiratory support and healthcare settings. Individual vaccination status, baseline vitamin C concentrations and product exposure were unavailable for informative treatment-selection analyses. The findings do not identify a deficient subgroup that should receive this high-dose regimen. Taken together, the randomised comparison, treatment delivery, sensitivity analyses and absence of a persuasive survival benefit support withholding routine intravenous vitamin C for COVID-19, while leaving nutritional replacement and other doses or indications as separate questions.^{1,4}

Body of Evidence

CITRIS-ALI supplied an important basis for the regimen. This blinded trial randomised 167 patients with sepsis and acute respiratory distress syndrome to vitamin C 50 mg/kg every six hours for 96 hours or placebo. Its primary modified organ failure score and inflammatory and vascular-injury biomarkers did not improve. Mortality at day 28 was 29.8% versus 46.3%, but this was one of 46 secondary outcomes and the exploratory comparisons were not adjusted for multiplicity. That mortality signal justified confirmation; it did not establish efficacy despite negative primary outcomes. LOVIT-COVID subsequently tested the same dose and duration in a much larger population with a primary endpoint incorporating survival.²

VITAMINS addressed the popular combination strategy. It randomised 216 patients with septic shock to vitamin C, hydrocortisone and thiamine or hydrocortisone alone. Vitamin C was given at 1.5 g every six hours, with treatment continued until shock resolution or for up to ten days. Among 211 evaluable patients, time alive and free of vasopressors through day 7 was 122.1 versus 124.6 hours; the reported median of paired differences was -0.6 hours (95% confidence interval -8.3 to 7.2; $P=0.83$). This did not support adding vitamin C and thiamine to hydrocortisone to accelerate shock resolution. Because both the dose and combined intervention differed, it could not settle the question of higher-dose vitamin C alone in viral respiratory failure.⁵

Direct early COVID-19 evidence was also limited. Zhang and colleagues conducted a blinded, placebo-controlled pilot in three Chinese hospitals, recruiting 56 critically ill patients before control of the outbreak curtailed enrolment. Vitamin C was given at 12 g every 12 hours for seven days. Median invasive ventilation-free days through day 28

were 26 versus 22, without a significant difference ($P=0.57$), and mortality was not significantly reduced. Better oxygenation and a lower interleukin-6 concentration were signals in secondary measurements. The small sample, incomplete recruitment and different regimen left substantial uncertainty. A physiological signal from this study could not establish the patient-centred benefit subsequently sought by the harmonised trials.⁶

LOVIT then provided a larger test of the same weight-based four-day regimen used in LOVIT-COVID. It randomised 872 adults with infection and vasopressor treatment early in their ICU admission. Death or persistent organ dysfunction at day 28 occurred in 191/429 (44.5%) vitamin C recipients and 167/434 (38.5%) placebo recipients, risk ratio 1.21 (95% confidence interval 1.04–1.40; $P=0.01$). Mortality alone was 35.4% versus 31.6%, risk ratio 1.17 (0.98–1.40). The composite established a concerning treatment effect, while the mortality component remained less precise. This was the external result that changed monitoring in the ongoing COVID-19 trials.³

The relationship between the studies is clinically informative. LOVIT concentrated on vasopressor-treated sepsis; LOVIT-COVID deliberately enrolled patients without preceding vasopressor use, and REMAP-CAP broadened the respiratory-failure and geographical spectrum. The harmonised result therefore extends concern beyond established circulatory shock. Its 28-day survival-without-dysfunction outcome also links the clinical question to LOVIT, despite a different primary analysis. The 63 COVID-19 participants from LOVIT were added only in a sensitivity analysis of the harmonised report and must not be counted as an independent third COVID-19 trial within that analysis.^{1,4}

By the end of 2023, the more persuasive randomised evidence did not support routine pharmacological vitamin C for either the studied sepsis population or hospitalised COVID-19. This synthesis does not require assuming that every dose, combination and stage of disease has the same effect. It reflects failure of the tested strategies to deliver reproducible patient benefit, with harm remaining a substantive concern. Any further treatment-selection hypothesis would need prospective evidence; unmeasured deficiency, isolated biomarkers and favourable exploratory periods do not supply a prescribing rule.^{1,3,4}

Should we use high-dose intravenous vitamin C in patients hospitalised with COVID-19?

Routine use of the studied four-day regimen is not supported: benefit was highly unlikely in both severity groups, with a concerning signal of harm.

References

1. The LOVIT-COVID Investigators, on behalf of the Canadian Critical Care Trials Group, and the REMAP-CAP Investigators. Intravenous Vitamin C for Patients Hospitalized With COVID-19: Two Harmonized Randomized Clinical Trials. *JAMA* 2023;330(18):1745–1759.
2. Fowler AA III, Truitt JD, Hite RD, Morris PE, DeWilde C, Priday A, et al. Effect of Vitamin C Infusion on Organ Failure and Biomarkers of Inflammation and Vascular Injury in Patients With Sepsis and Severe Acute Respiratory Failure: The CITRIS-ALI Randomized Clinical Trial. *JAMA* 2019;322(13):1261–1270.
3. Lamontagne F, Masse MH, Menard J, Sprague S, Pinto R, Heyland DK, et al; LOVIT Investigators and the Canadian Critical Care Trials Group. Intravenous Vitamin C in Adults with Sepsis in the Intensive Care Unit. *N Engl J Med* 2022;386(25):2387–2398.
4. The LOVIT-COVID Investigators and the REMAP-CAP Investigators. Supplement 1 to: Intravenous Vitamin C for Patients Hospitalized With COVID-19: Two Harmonized Randomized Clinical Trials. *JAMA* 2023;330:1745–1759.
5. Fujii T, Luethi N, Young PJ, Frei DR, Eastwood GM, French CJ, et al; VITAMINS Trial Investigators. Effect of Vitamin C, Hydrocortisone, and Thiamine vs Hydrocortisone Alone on Time Alive and Free of Vasopressor Support Among Patients With Septic Shock: The VITAMINS Randomized Clinical Trial. *JAMA* 2020;323(5):423–431.
6. Zhang J, Rao X, Li Y, Zhu Y, Liu F, Guo G, et al. Pilot trial of high-dose vitamin C in critically ill COVID-19 patients. *Ann Intensive Care* 2021;11:5.

CONFIDENT

Misset B, Piagnerelli M, Hoste E, Dardenne N, Grimaldi D, Michaux I, et al. Convalescent Plasma for Covid-19–Induced ARDS in Mechanically Ventilated Patients. N Engl J Med 2023;389(17):1590–1600.

Introduction

Convalescent plasma offered an immediately available approach to passive immunisation during the COVID-19 pandemic. Antibodies collected from recovered donors could potentially neutralise severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) before recipients generated an effective response. For patients already requiring invasive ventilation, however, a plausible antiviral intervention faced a difficult question: was continuing viral replication still a sufficiently important driver of illness for additional antibodies to improve survival? Established lung injury and the host inflammatory response might limit the value of treatment at this stage.¹

The product also mattered. Antibody binding and functional neutralisation are different measurements, and plasma collected from different donors is not a uniform drug. A biologically credible strategy required attention to antibody activity, administered volume and the virus against which activity had been measured. Transfusion added fluid and exposure to a blood product in patients with acute respiratory distress syndrome (ARDS), so even a readily available treatment required evidence of a favourable clinical balance.¹

The evidence changed during CONFIDENT's recruitment. A small trial published in early 2021 supported plasma administered within three days of mild symptoms in older adults. Subsequently, RECOVERY and REMAP-CAP failed to establish benefit in hospitalised and critically ill populations. These results made it increasingly important to define what an additional trial could resolve, rather than regard all plasma products and treatment windows as interchangeable.^{2,3,4}

CONFIDENT concentrated on patients with COVID-19–induced ARDS receiving invasive ventilation and used plasma selected by a functional neutralisation assay. It encouraged enrolment soon after intubation and stratified allocation at 48 hours. The hypothesis was that sufficiently active plasma, delivered early in the course of invasive support, would reduce 28-day mortality. This was a specific treatment strategy within established critical illness, whose timing required distinction from treatment early after infection.¹

Synopsis

CONFIDENT was an investigator-led, randomised, open-label, parallel-group superiority trial conducted at 17 Belgian sites and coordinated by the University Hospital of Liège. The Belgian Health Care Knowledge Centre funded the study. The university hospital was the sponsor and owned the data; investigators designed the trial, conducted its analyses and prepared the report. An independent data and safety monitoring board reviewed interim results, which remained concealed from investigators and the steering committee.^{1,5}

Adults were eligible with COVID-19–induced ARDS, a Clinical Frailty Scale score below 6 and invasive mechanical ventilation for no more than five days. ARDS used the Berlin definition. The principal report required interstitial pneumonia on chest imaging within ten days and a positive nasopharyngeal SARS-CoV-2 polymerase-chain-reaction test within 15 days before inclusion. Pregnancy, previous transfusion-related adverse effects, an existing decision to limit treatment and participation in another COVID-19 trial were exclusions. The coordinating centre checked eligibility and investigators confirmed the availability of ABO-identical plasma before seeking consent. Relatives initially consented for every enrolled patient.¹

A web-based system allocated patients equally to plasma plus standard care or standard care alone. Randomisation was stratified by time from starting invasive ventilation: 48 hours or less, or more than 48 through 120 hours. It was not stratified by centre. Investigators and treating teams knew allocation; there was no placebo. The published methods describe the allocation system without supplying sufficient detail to reconstruct its sequence-generation or concealment procedures.^{1,5}

The intervention comprised two units of ABO-identical convalescent plasma, totalling 400–500 ml, to be infused within 24 hours of inclusion. Donors had recovered from COVID-19 between 28 days and ten months previously. Plasma was tested against a Wuhan-like SARS-CoV-2 isolate using a live-virus assay. The requested 50% neutralisation titre, or NT₅₀, was at least 1:320; 1:160 was accepted when supplies were limited. This titre expresses activity at a specified dilution, rather than the total antibody quantity delivered. Crossover was prohibited.¹

Both groups received supportive ARDS care. Published recommendations included restrictive fluid management, tidal volumes around 6 ml/kg predicted body weight, plateau pressure no greater than 30 cmH₂O and individualised positive end-expiratory pressure. Proning and early neuromuscular blockade were recommended for more severe hypoxaemia, with extracorporeal support available as rescue. Glucocorticoids and

thromboprophylaxis formed part of standard care. These were recommendations accompanying the intervention, not a separately randomised ventilation strategy.⁵

The primary endpoint was death by day 28 after randomisation. Secondary outcomes included adverse events, inflammatory and antibody responses, Sequential Organ Failure Assessment (SOFA) scores, organ support, hospital stay and mortality at days 90 and 365. Time alive and free of support was assessed through day 28. The primary analysis used intention to treat, a chi-square comparison and logistic regression adjusting for the randomisation stratum. No primary outcome was missing. Secondary outcomes were analysed without imputation; their confidence intervals were not adjusted for multiplicity and the analyses were exploratory.¹

The sample-size calculation assumed 40% control mortality and sought approximately a one-third relative reduction, specified as a 13.5-percentage-point absolute difference. With two-sided alpha 0.05 and 80% power, 500 patients were required. Sequential monitoring used Lan-DeMets spending functions, with published futility and efficacy boundaries at 100, 200, 300 and 400 participants and a planned final assessment at 500. The latter listed an efficacy threshold of $P < 0.042$. The monitoring board recommended continuation after each of four reviews.^{1,5}

The principal report dates recruitment from 10 September 2020 to 9 March 2022. Of 1,034 screened patients, 475 were enrolled: 237 assigned plasma and 238 standard care. Common reasons for non-enrolment included timing criteria, another trial, treatment limitations and refusal. Recruitment ended on 9 April 2022 after a month without a newly screened patient, leaving 25 of the planned 500 unfilled. There were 171 patients per group in the earlier ventilation stratum and 66 versus 67 in the later stratum. Four patients declined continued participation after recovery, but vital-status collection was retained.^{1,5}

Median age was 64 years in both groups, and approximately two-thirds were men. Median symptom duration was 12 days, hospital stay before randomisation five days and intensive care stay three days. Median SOFA score was 6; the median ratio of arterial oxygen tension to inspired oxygen fraction was 117 versus 128 mmHg. Moderate or severe ARDS accounted for 90.1% of participants. Documented vaccination was uncommon, although collection of vaccination information varied by centre. Hypertension was more frequent in the plasma group and diabetes in controls.¹

All but one assigned patient received plasma; that patient died before infusion. No control received it. The main report states that 17.7% received plasma with a titre of 1:160, with the remainder receiving higher-titre product. Glucocorticoids were given to

233 patients in each group, 98.1% overall. Remdesivir use was 5.5% versus 5.9%, and interleukin-6 pathway treatment 5.1% versus 2.9%. Recorded tidal volumes and pressures were broadly similar between groups, although some ventilation measurements were missing.^{1,5}

By day 28, 84/237 plasma recipients and 107/238 controls had died, 35.4% versus 45.0%; $P=0.03$ both before and after stratification adjustment. Independently calculated from these counts, the unadjusted risk difference was -9.5 percentage points, with an approximate Wald 95% confidence interval of -18.3 to -0.7 . This descriptive calculation is not a stratified or sequentially adjusted estimate. Deaths in the earlier stratum were 56/171 versus 80/171, 32.7% versus 46.8%; in the later stratum they were 28/66 versus 27/67, 42.4% versus 40.3%. Survival curves separated at approximately day 17.¹

Day-90 mortality was 102/237 versus 121/238, 43.0% versus 50.8%, an absolute difference of -7.8 percentage points (95% CI -16.8 to 1.2). Median ventilation-free days through day 28 were zero in both groups; ventilation duration and lengths of stay showed no clear separation. At least one adverse event affected 158 versus 173 patients. There were 93 versus 116 serious adverse events, including 78 versus 106 fatal events; these are event counts. None was attributed to plasma. Decisions to limit treatment before day 28 occurred in 21 versus 32 patients, 8.9% versus 13.4%.¹

Critique

CONFIDENT provides a positive randomised mortality comparison in a clinically coherent population. Its most consequential uncertainty is whether this establishes a reproducible effect of the tested strategy, and particularly whether the apparent concentration of benefit within 48 hours defines a reliable treatment rule. The overall result deserves attention: mortality was the primary endpoint, follow-up for it was complete and the exposure contrast was almost fully maintained. Those protections distinguish the finding from an association among patients selected to receive plasma. They do not, however, settle the timing hypothesis or explain discordance with previous trials.¹

The point estimate is clinically substantial, but uncertainty remains wide. The independently calculated crude interval spans a small absolute benefit to a much larger one, and the paper does not supply an adjusted effect estimate with its confidence interval for the primary logistic model. The P value of 0.03 is below the tabulated planned final threshold of 0.042, although exact spending at the achieved sample of 475 is not shown in the available report. This is a reporting limitation, not grounds to label the result automatically invalid. Recruitment ceased because patients were no longer being screened, not because a favourable interim boundary had been crossed.^{1,5}

Stratifying randomisation at 48 hours gives the timing comparison greater credibility than an entirely retrospective division. Nevertheless, the trial randomised plasma within each stratum; it did not randomise patients to early or late treatment. The earlier stratum contained 342 patients, compared with only 133 later, and neither the principal report nor supplied results provides a treatment-by-timing interaction estimate with uncertainty. Different within-stratum results do not themselves prove that effects differ. The later estimate is too imprecise to establish that plasma becomes ineffective after 48 hours, while the earlier result remains vulnerable to chance subgroup variation.^{1,5}

Moreover, early after intubation was typically late after symptom onset. The median 12-day illness duration makes it inappropriate to import the rationale for very early antiviral therapy without qualification. Intubation timing also depends on preceding respiratory support, transfer and clinical practice. A 48-hour ventilation threshold may therefore classify different biological stages across hospitals. The trial's timing signal is a reason to test a more precisely defined strategy; it does not establish that intubation starts a universal therapeutic clock.¹

Death is an objective event, but open treatment allocation can influence the care preceding it. Decisions to limit therapy were more frequent among controls, with reported median decision times of 13 rather than 17 days. These decisions may reflect evolving illness, and a treatment effect could itself alter their frequency. Equally, knowledge of receiving an experimental intervention could influence persistence with support. The direction and magnitude of any such influence cannot be established here. Similar recorded ventilation settings provide some reassurance but do not capture every treatment decision, and plateau-pressure data were missing for many patients.^{1,5}

Centre variation adds a concrete concern. At one substantial recruiting centre, mortality was 12/40 with plasma and 5/38 with standard care, reversing the overall direction. Randomisation was not stratified by site, and small numbers elsewhere limited assessment of treatment-by-centre interaction. The authors proposed differences in time from intensive care admission to enrolment as an explanation, but the supplementary timing table does not consistently support the direction stated in the main report. This unresolved explanation should not be used to dismiss an inconvenient centre result. Nor would removing that centre after seeing outcomes provide an unbiased estimate.^{1,5}

The functional assay and near-complete treatment delivery are strengths, but they do not define a universally reproducible plasma product. Nearly one in five recipients received the lower accepted titre. Plasma potency was measured against an ancestral isolate, while recruitment continued through subsequent variant periods. There was no

calibration to an international standard, and the product-titre mortality analysis used the higher titre of the two administered bags. Neither a high nominal titre nor the absence of an observed titre–mortality association proves an adequate dose against each recipient’s virus. Titre groups were not randomly assigned.^{1,5}

Mechanistic and subgroup findings provide limited additional certainty. Recipient antibody concentrations increased in both groups, without a clear treatment-specific pattern establishing the proposed pathway. Baseline neutralisation measurements were available for 346 of 475 patients. Sampling depended partly on centre participation and laboratory success; these administrative explanations do not establish that missingness was unrelated to prognosis or response. Variant subgroups were assigned by the predominant strain in Belgium rather than individual sequencing, and only 16 patients came from the Omicron period. Assertions of similar effects across antibodies or variants therefore exceed what these exploratory comparisons can reliably exclude.⁵

Secondary outcomes do not independently reproduce the mortality finding. The day-90 estimate remained favourable but its interval included no difference. The reported one-year estimate likewise crossed the null, and its reduced denominators do not reconcile straightforwardly with the statement of complete vital-status follow-up. Neither result proves that the 28-day effect disappeared, but sustained survival benefit is less secure. A lower inflammatory marker among day-seven survivors is affected by conditioning on survival and by multiple comparisons. Zero median ventilation-free days in both groups also illustrates how a median can obscure changes elsewhere in a distribution.¹

Safety should be expressed with similar restraint. No serious event was judged directly attributable to plasma, but attribution was unblinded and 236 treated patients cannot exclude uncommon transfusion complications. Many recorded serious events were fatal manifestations of the underlying illness; their lower count is not a separate demonstration of product safety. The trial supports feasibility under its blood-service and critical care arrangements, with no identified direct safety signal, rather than equivalence in harm.¹

Applicability is strongest to the enrolled Belgian population: invasively ventilated, mostly moderately or severely hypoxaemic, without marked frailty or a pre-existing treatment limitation, and almost universally receiving glucocorticoids. Antiviral and additional immunomodulatory exposure was limited, and documented vaccination was infrequent. Rapid access to compatible, functionally tested plasma was integral to enrolment. Extrapolating the effect to different immunity, variants, co-treatment or blood-service capability requires evidence. Taken together, the trial supports a

potentially important benefit requiring confirmation, with the early-ventilation subgroup providing a focused hypothesis rather than a settled indication.^{1,5}

Body of Evidence

The preceding trials address several distinct questions: preventing deterioration during mild infection, treating established hospitalised disease and reversing organ failure in critical illness. Their differences help interpret CONFIDENT, but should not be used to explain away every inconsistent result. Timing and antibody activity are plausible effect modifiers; establishing either requires more than comparing which trials reached statistical significance.

Libster and colleagues reported a double-blind trial in 160 older adults with mild COVID-19, administering high-titre plasma or saline within 72 hours of symptom onset. Severe respiratory disease developed in 13/80 versus 25/80 patients, 16% versus 31%; relative risk 0.52 (95% CI 0.29–0.94). The endpoint used tachypnoea or reduced oxygen saturation, rather than death. Recruitment stopped at 76% of the intended sample as local cases declined. This result supported intervention early in infection but supplied little direct evidence for treating ventilated ARDS. The shared description of “early” treatment conceals markedly different illness stages in Libster and CONFIDENT.²

RECOVERY then provided a much larger hospital-based comparison of high-titre plasma plus usual care against usual care alone. Among 11,558 randomised patients, 28-day deaths were 1,399/5,795 and 1,408/5,763, both approximately 24%; the reported mortality rate ratio was 1.00 (95% CI 0.93–1.07). Results were similar in the prespecified subgroup without detectable baseline SARS-CoV-2 antibodies. Progression to invasive ventilation or death among those not already ventilated was also unchanged. The size and precision of this result argue strongly against a large average benefit in the broader hospital population. They do not directly resolve CONFIDENT’s selected ventilated population, but they make a universal plasma effect implausible.³

REMAP-CAP supplied a closer critical care comparison, enrolling 2,011 critically ill participants in its immunoglobulin domain. Two units of high-titre, ABO-compatible plasma were compared with no plasma. The primary ordinal outcome combined hospital death with days free of respiratory or cardiovascular support through day 21. The adjusted odds ratio was 0.97 (95% credible interval 0.83–1.15), with a 99.4% probability of futility, defined against an odds ratio of 1.2. Hospital mortality was 37.3% versus 38.4%. Approximately one-third were invasively ventilated at baseline; the population therefore overlapped with, but was broader than, CONFIDENT. Differences in endpoint

and respiratory support matter, yet this substantial negative critical care trial remains an important counterweight.⁴

The Belgian DAWn-plasma trial is particularly relevant to product potency. It randomised 483 hospitalised patients in a 2:1 ratio to four plasma units plus standard care or standard care alone. Median symptom duration at randomisation was seven days, and median infused volume was 884 ml; approximately 81% of units had an NT₅₀ of at least 1:320. At day 15, 83.74% versus 84.05% were alive and free of mechanical ventilation, odds ratio 0.99 (95% CI 0.59–1.66). Antibody trajectories were not altered. Although its clinical question differed, use of a substantial volume selected for high functional titres did not reliably produce benefit. Potency alone therefore cannot reconcile CONFIDENT with the preceding evidence.⁶

CONFIDENT adds a positive mortality trial restricted to invasively ventilated ARDS and preserves a credible randomised comparison despite difficult pandemic recruitment. It challenges the assumption that plasma must be ineffective once invasive support has begun. Its distinctive features nevertheless arrived together: selection by severity and frailty, an early ventilation window, a particular neutralisation assay, locally available plasma and a particular background of care. The trial cannot isolate which, if any, explains its result. A favourable finding within one combination does not validate each component as a treatment-selection rule.¹

By the end of 2023, the evidence therefore supported caution about routine use while retaining a testable question in this selected population. Confirmation would need to preserve the mortality comparison while documenting product activity against the relevant virus, recipient immunity, timing and treatment-limitation decisions. A trial designed to assess effect modification could also determine whether the 48-hour signal is reproducible. Until then, CONFIDENT is an important challenge to the larger negative evidence base, with insufficient certainty to make early post-intubation plasma a standard treatment.

Should we use convalescent plasma routinely for invasively ventilated COVID-19 ARDS?

Not routinely. CONFIDENT's mortality benefit warrants confirmation, particularly before adopting its apparent early-ventilation benefit as a treatment rule.

References

1. Misset B, Piagnerelli M, Hoste E, Dardenne N, Grimaldi D, Michaux I, et al. Convalescent Plasma for Covid-19–Induced ARDS in Mechanically Ventilated Patients. *N Engl J Med* 2023;389(17):1590–1600.
2. Libster R, Pérez Marc G, Wappner D, Coviello S, Bianchi A, Braem V, et al; Fundación INFANT–COVID-19 Group. Early High-Titer Plasma Therapy to Prevent Severe Covid-19 in Older Adults. *N Engl J Med* 2021;384(7):610–618.
3. RECOVERY Collaborative Group. Convalescent plasma in patients admitted to hospital with COVID-19 (RECOVERY): a randomised controlled, open-label, platform trial. *Lancet* 2021;397(10289):2049–2059.
4. Writing Committee for the REMAP-CAP Investigators. Effect of Convalescent Plasma on Organ Support–Free Days in Critically Ill Patients With COVID-19: A Randomized Clinical Trial. *JAMA* 2021;326(17):1690–1702.
5. Misset B, Piagnerelli M, Hoste E, Dardenne N, Grimaldi D, Michaux I, et al. Supplementary Appendix to: Convalescent Plasma for Covid-19–Induced ARDS in Mechanically Ventilated Patients. *N Engl J Med* 2023;389:1590–1600.
6. Devos T, Van Thillo Q, Compernelle V, Najdovski T, Romano M, Dauby N, et al. Early high antibody titre convalescent plasma for hospitalised COVID-19 patients: DAWn-plasma. *Eur Respir J* 2022;59(2):2101724.

Possible Reporting Inconsistencies. The proposed timing explanation for centre 105 is inconsistent across the report. The Results and Figure 2 caption say standard-care patients entered sooner after ICU admission, but the table accompanying supplementary Figure S8 gives a median of three days with plasma and four with standard care. The stated explanation for this centre’s reversed treatment effect is therefore uncertain.^{1,5}

STRESS-L

Whitehouse T, Hossain A, Perkins GD, Gordon AC, Bion J, Young D, et al; STRESS-L Collaborators. Landiolol and Organ Failure in Patients With Septic Shock: The STRESS-L Randomized Clinical Trial. JAMA 2023;330(17):1641–1652.

Introduction

Persistent tachycardia during septic shock presents an uncomfortable therapeutic dilemma. A faster heart rate may sustain cardiac output when stroke volume is limited, but prolonged adrenergic stimulation may also increase myocardial oxygen demand and impair ventricular filling. Correcting the underlying infection, hypovolaemia, pain or inadequate perfusion remains essential. Once these problems have been addressed, it is less clear whether the residual heart rate is a useful compensatory response or a potentially modifiable contributor to organ injury. A heart-rate target alone cannot distinguish those possibilities.¹

Short-acting β -blockade offered an attractive way to test this hypothesis. An intravenous infusion could be titrated, monitored and stopped quickly if circulation deteriorated. However, rapid reversibility does not remove the consequences of reducing cardiac output in a patient dependent on sympathetic support. A treatment might achieve its heart-rate target while increasing vasopressor requirements or worsening tissue perfusion. The relevant question is therefore whether suppressing tachycardia improves the patient's course, beyond changing the monitor display.¹

A single-centre esmolol trial had demonstrated effective rate control and an unexpectedly large secondary mortality difference favouring treatment. Its closely monitored haemodynamic strategy encouraged further investigation, but did not establish that a similar approach would work across routine intensive care units or with another drug.² Landiolol, an ultra-short-acting β_1 -selective antagonist, was selected for STRESS-L because its pharmacological properties appeared suitable for cautious titration during shock.¹

STRESS-L, the Study into the Reversal of Septic Shock with Landiolol, moved the question from rate control towards organ failure. It examined patients whose tachycardia and substantial norepinephrine requirement persisted after at least a day of treatment. The hypothesis was that adding landiolol to standard care would reduce the burden of organ dysfunction over the next two weeks. This was a test of deliberately initiating β -blockade

in established septic shock, with both clinical benefit and haemodynamic harm remaining plausible.¹

Synopsis

STRESS-L was an investigator-initiated, phase 2b, open-label, parallel-group randomised trial involving 40 UK National Health Service hospitals. Recruitment ran from 19 April 2018 to 15 December 2021, with a pause during the first pandemic wave. The National Institute for Health and Care Research funded the study, University Hospitals Birmingham sponsored it and AOP Health supplied landiolol without charge. The funder and drug supplier had no role in conducting the study, analysing the data, writing the report or deciding to publish; the funder approved the protocol.¹

Eligible patients were adults with septic shock, a heart rate of at least 95 beats/min and a norepinephrine infusion of at least 0.1 µg/kg/min for more than 24 but less than 72 hours. Infection, an increase in Sequential Organ Failure Assessment (SOFA) score of at least two points and lactate above 2 mmol/l at some point during resuscitation supported the sepsis definition. Adequate fluid resuscitation was judged clinically. Important exclusions included tachycardia attributed to pain or discomfort, non-infectious vasodilatory shock, substantial hepatic impairment and an expected survival below 28 days. Consent was obtained from patients or their legal representatives. Allocation was 1:1; the available clinical report does not provide sufficient technical detail to independently assess sequence generation or concealment. Treatment allocation was known to clinicians and investigators, although investigators remained blinded to aggregate group data during recruitment.¹

Landiolol started intravenously at 1 µg/kg/min and increased by 1 µg/kg/min every 15 minutes towards a heart rate of 80–94 beats/min, expected within six hours. Published supplementary treatment instructions allowed rates up to 40 µg/kg/min. Infusion was reduced when heart rate fell below 80 beats/min, and treatment could continue during vasopressor support for up to 14 days. After vasopressors had been stopped for 12 hours, landiolol was weaned gradually. It was to be stopped for at least 12 hours before ICU discharge and not restarted after a 12-hour interruption.^{1,3}

Both groups received infection treatment, source control, appropriate antimicrobials and fluid resuscitation. The intended control strategy excluded β-blockers during the ICU stay, although supplementary deviation records document control patients receiving them. Blood-pressure support was managed separately from heart-rate titration, aiming for a mean arterial pressure of at least 65 mm Hg or a predefined individual target. Cardiac output monitoring, echocardiography and positive inotropes remained

discretionary. Consequently, the intervention was a rate-targeted drug strategy within usual critical care, without mandatory measurement of cardiac output.^{1,3}

The primary outcome was the mean modified SOFA score while alive in ICU over 14 days after randomisation. Neurological assessment was omitted because of sedation, leaving five organ components and a total score from zero to 20, with lower scores favourable. Each patient's daily scores were summed and divided by the number of observed days; death and ICU discharge ended observation. Secondary outcomes included 28- and 90-day mortality, length of stay among survivors, vasopressor use, heart rate, blood pressure and biochemical measures. Serious adverse events were followed to day 90.¹

The intended sample was 340 patients. The effect-size and power assumptions underlying that calculation are not specified in the available clinical report. The primary analysis included all randomised patients and used linear mixed modelling with site as a random effect; adjusted estimates included age, sex and baseline norepinephrine dose. Confidence intervals used 10,000 bootstrap replications. Missing data were imputed for the primary outcome, and alternative imputation and complier analyses were presented as sensitivity analyses. The independent monitoring committee reviewed unblinded data without formal statistical stopping boundaries. Recruitment stopped at 126 patients because benefit was considered unlikely and a possible harm signal had emerged.^{1,3}

Of 4,137 screened patients, 348 were eligible and 126 were randomised, 63 to each group. Mean age was 55.6 years and 58.7% were male. Pulmonary and abdominal infections predominated. Baseline mean heart rates were 110.6 and 114.1 beats/min, norepinephrine doses 0.37 and 0.36 µg/kg/min, and modified SOFA scores 10.1 and 10.3, respectively. Atrial fibrillation was present in seven and eight patients, so eligibility was not confined to sinus tachycardia. All 126 contributed to the primary analysis; one landiolol patient lacked mortality follow-up.¹

Sixty patients received landiolol, for a median of 2.0 days. Nine patients stopped assigned treatment while remaining in study follow-up. Mean heart rate over 14 days was 92.4 versus 98.6 beats/min, an unadjusted modelled difference of -6.46 beats/min (95% CI -10.46 to -2.46; $P = 0.002$). Thus, a physiological treatment contrast was achieved, although controls also improved. Five landiolol patients were classified as non-compliers, and supplementary records describe additional dosing, titration and concomitant-treatment deviations.^{1,3}

Mean modified SOFA scores were 8.8 with landiolol and 8.1 with standard care. The unadjusted mean difference was 0.75 points (95% CI -0.49 to 2.00; $P = 0.24$), and the adjusted difference was 0.63 (95% CI -0.47 to 1.73; $P = 0.26$). The median number of days

contributing to this outcome was seven versus 11. Sensitivity analyses did not demonstrate benefit. There was no convincing interaction with baseline norepinephrine requirement; the interaction P value was 0.47.^{1,3}

At 28 days, 23/62 landiolol patients and 16/63 controls had died, 37.1% versus 25.4%. The unadjusted absolute difference was 11.7 percentage points (95% CI -4.4 to 27.8; P = 0.16). At 90 days, deaths numbered 27/62 and 18/63, 43.5% versus 28.6%, a difference of 15.0 percentage points (95% CI -1.7 to 31.6; P = 0.08). These mortality comparisons were imprecise and did not reach conventional statistical significance.¹

Mean arterial pressure over five days was 73.2 versus 76.0 mm Hg, an unadjusted difference of -2.67 mm Hg (95% CI -5.06 to -0.29; P = 0.03). Norepinephrine requirements were numerically greater with landiolol: mean rates were 0.34 versus 0.24 µg/kg/min, with an adjusted difference of 0.07 µg/kg/min (95% CI -0.003 to 0.15; P = 0.06). Lactate was also numerically higher, without a statistically conclusive difference. Serious adverse events affected 16/63 patients versus 4/63, 25.4% versus 6.4% (P = 0.006). These are numbers of affected patients, rather than total events or a count of events proved to be caused by landiolol.^{1,3}

Critique

The main interpretative issue is the discordance between effective heart-rate reduction and the absence of clinical benefit, accompanied by several concerning safety observations. STRESS-L does not establish a statistically certain increase in mortality. Equally, its non-significant mortality comparisons do not establish safety: the confidence intervals include substantial absolute harm. Against that uncertainty, neither organ failure nor haemodynamic support provided a compensating signal of benefit. Routine implementation requires evidence that its advantages justify its risks; that requirement was not met.¹

The primary outcome complicates interpretation because it averages organ dysfunction only while patients remain alive in ICU. Both rapid recovery and early death shorten the measurement period, although their clinical meanings are opposite. A patient dying after several severely abnormal days can no longer accumulate later observations, while a surviving control may contribute a longer trajectory. The seven- versus 11-day median observation periods show that this was a material feature of the data. Mean SOFA is therefore not a complete measure of survival with recovered organ function. Its numerical direction should be considered alongside mortality, rather than treated as the sole verdict on net benefit.¹

There is a second connection between outcome and treatment: the cardiovascular SOFA component incorporates vasopressor support. A strategy that lowers heart rate but requires more norepinephrine can worsen that component without proving new structural organ injury. This is not simply an unfair endpoint. Greater dependence on cardiovascular support is itself relevant when assessing a drug intended to improve shock. However, the composite cannot tell whether landiolol impaired intrinsic cardiac performance, altered treatment decisions or affected other organs. The absence of mandatory cardiac output measurements limits the mechanistic explanation, even though it does not invalidate the randomised comparison.¹

Early stopping substantially reduces precision. Only 37% of the intended sample was enrolled, and there were no formal stopping boundaries defining a numerical threshold for harm or futility. The decision incorporated mortality, physiology and the low prospect of demonstrating benefit. Such a decision can be clinically reasonable, but observed differences at an early stop may overstate the eventual treatment effect. It would therefore be inappropriate to convert the 90-day absolute difference into a stable number needed to harm. Conversely, continuing a trial solely to narrow confidence intervals would require adequate equipoise in the face of these observations.¹

The safety pattern deserves more attention than a simple division into significant and non-significant P values. Lower arterial pressure, greater norepinephrine exposure, numerically higher lactate, more serious adverse events and more deaths all point towards concern, although they are correlated observations rather than independent confirmations. They are compatible with impaired circulatory tolerance, but do not demonstrate that mechanism. Multiple secondary analyses also make isolated probability values less decisive. The appropriate conclusion is a coherent possible harm signal, with uncertainty about its size and cause, rather than proof that each adverse physiological change was treatment-induced.^{1,3}

Open treatment assignment could influence titration, the use of additional drugs and adverse-event attribution. Norepinephrine dose and ICU length of stay are partly determined by clinical decisions; serious-event reporting may be more attentive around an unfamiliar infusion. Mortality is less vulnerable to ascertainment bias, and follow-up was nearly complete, which makes its direction important despite the limited precision. Randomisation and broadly similar baseline severity protect against many confounding explanations. These strengths support taking the findings seriously without assuming that every observed difference represents a direct pharmacological effect.¹

Implementation was demanding. The supplement records failures to start treatment, titration outside the intended heart-rate range, dosing errors and β -blocker

administration in control patients. Some individual descriptions link hypotension temporally to landiolol and improvement to its withdrawal, but such narratives cannot estimate a causal event rate. Contamination may have reduced the difference between strategies, while dosing problems may have increased risk. Neither can be dismissed as irrelevant to adopting a tightly titrated infusion across ordinary ICUs. The practical treatment effect includes the reliability with which a strategy can be delivered.³

Removing patients who became unstable or could not tolerate treatment would not necessarily recover the drug's true benefit. Tolerance is determined after randomisation and may itself reflect the intervention. A per-protocol comparison can therefore preferentially retain patients with greater cardiovascular reserve. The supplementary complier analysis did not establish benefit, and its causal interpretation requires assumptions beyond randomisation alone. Intention-to-treat remains the appropriate starting point for deciding whether to introduce this treatment strategy.^{1,3}

Selection also matters. Patients needed persistent tachycardia and substantial norepinephrine support after more than 24 hours of shock treatment. This excludes a large fraction of screened patients and does not describe initial resuscitation, late prolonged shock or a patient with an isolated arrhythmia requiring rate control. The requirement for adequate fluid resuscitation was clinically judged rather than linked to a mandatory cardiac assessment. A heart rate above 95 beats/min is an observable threshold, but does not identify whether reducing that rate will preserve cardiac output. The trial consequently tests a recognisable clinical phenotype without validating it as a responsive biological subgroup.¹

The multicentre setting improves relevance to UK practice, but recruitment was sparse and difficult. Pandemic interruption added delay, yet only one participant had COVID-19; this was not principally a COVID-19 shock study. The incomplete sample and small subgroups provide little support for claims that benefit survives in a particular norepinephrine stratum or respiratory phenotype. An interaction P value above 0.05 does not prove identical effects across those groups. Nor can this study settle the separate question of continuing established long-term β -blockade. Its most defensible contribution is to discourage routine initiation of landiolol solely to normalise persistent tachycardia in this population.^{1,3}

Body of Evidence

The enthusiasm preceding STRESS-L arose largely from Morelli and colleagues' 2013 trial. In 154 patients, esmolol targeted 80–94 beats/min after haemodynamic optimisation, with pulmonary artery catheter monitoring and access to levosimendan

rescue. Rate control, the primary objective, was achieved. The striking secondary finding was 28-day mortality of 49.4% versus 80.5%, with an adjusted hazard ratio of 0.39 (95% CI 0.26–0.59). This single-centre, open-label result supplied a powerful hypothesis, but its unusually high control mortality and closely supervised treatment package limited straightforward generalisation. It was evidence to justify replication, not a secure estimate of the survival benefit of β -blockade throughout septic shock.²

STRESS-L was not an exact replication. It used another drug, a broader multicentre delivery model and no mandatory cardiac output monitoring. These differences can generate plausible explanations for divergent results, but none was randomised between the two studies. It is not possible to assign the discrepancy specifically to landiolol, monitoring, rescue inotropes or differences in recruited patients. Claiming that one missing element would have restored benefit would require a further comparison. What STRESS-L directly contributes is evidence about its own treatment strategy under the conditions in which clinicians would need to implement it.^{1,2}

The Japanese J-Land 3S trial addressed a related but narrower physiological objective. It randomised 151 patients with sepsis-related tachyarrhythmia, including sinus tachycardia, atrial fibrillation and flutter. At 24 hours, 55% versus 33% of the analysed patients reached 60–94 beats/min, a reported difference of 23.1 percentage points (95% CI 7.1–37.5; $P = 0.0031$). This supports landiolol's ability to control rate. However, achieving that endpoint cannot establish that suppressing tachycardia improves organ function or survival. Drug-related serious events included blood-pressure reduction, reinforcing the need to distinguish successful rate control from safe circulatory support. Its entry conditions and primary outcome do not make it a contradictory positive organ-failure trial.⁴

A smaller trial by Cocchi and colleagues provided another caution before the end of 2023. Forty-two patients were randomised and 40 received study interventions, including 18 assigned esmolol and 22 usual care. The primary comparison was norepinephrine-equivalent dose six hours after treatment initiation. Mean doses were 0.30 versus 0.21 $\mu\text{g/kg/min}$ ($P = 0.15$), providing no evidence of the anticipated reduction in vasopressor requirement. The small sample and short physiological endpoint preclude a firm mortality conclusion. Nevertheless, it weakened the expectation that reducing adrenergic drive would reliably make vasopressor withdrawal easier. It also demonstrates why the earlier survival signal should not be treated as a consistent class effect.⁵

Evidence about continuing chronic β -blockers asks a different question. Fuchs and colleagues analysed 296 patients already receiving long-term treatment; therapy

continued in 167 and stopped in 129. Continuation was associated with lower 90-day mortality, 40.7% versus 52.7%, with an adjusted hazard ratio of 0.67 (95% CI 0.48–0.95). This was an observational secondary analysis. Clinicians' ability to continue medication may reflect haemodynamic stability and other prognostic features that adjustment cannot fully remove. It neither proves benefit from starting an infusion during shock nor establishes that chronic treatment should always be withdrawn. The distinction matters because STRESS-L cannot answer every clinical question involving a β -blocker and sepsis.⁶

By the CCR24 evidence cut-off, the studies therefore supported a more discriminating interpretation than a single pooled proposition that β -blockade is beneficial. The drugs could reduce heart rate, but trials differed in haemodynamic preparation, monitoring, rhythm, treatment timing and chosen outcome. A positive physiological endpoint and a hypothesis-generating secondary survival result did not outweigh a multicentre organ-failure trial with no demonstrated benefit and possible harm. Future investigation would need to identify patients whose cardiac output tolerates rate reduction and then demonstrate patient-important benefit. STRESS-L supplies no validated bedside selection rule that permits assuming that benefit already exists.^{1,4,5}

Should landiolol be started to control persistent tachycardia in established septic shock?

No. It lowered heart rate without improving organ failure and produced a concerning signal of harm.

References

1. Whitehouse T, Hossain A, Perkins GD, Gordon AC, Bion J, Young D, et al; STRESS-L Collaborators. Landiolol and Organ Failure in Patients With Septic Shock: The STRESS-L Randomized Clinical Trial. *JAMA* 2023;330(17):1641–1652.
2. 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.
3. Whitehouse T, Hossain A, Perkins GD, et al. Supplement 3 to: Landiolol and Organ Failure in Patients With Septic Shock: The STRESS-L Randomized Clinical Trial. *JAMA*. 2023.

4. Kakihana 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. Fuchs C, Wauschkuhn S, Scheer C, Vollmer M, Meissner K, Kuhn S-O, et al. Continuing chronic beta-blockade in the acute phase of severe sepsis and septic shock is associated with decreased mortality rates up to 90 days. *Br J Anaesth* 2017;119(4):616–625.

Possible Reporting Inconsistencies. Table 2 appears to have a missing minus sign before the lower confidence limit for the adjusted 90-day mortality risk difference: it prints 12.77 percentage points, with a 95% confidence interval of 2.00 to 27.54 and $P=0.09$. The interval should be checked against the original analysis before reuse.¹

ACORN

Qian ET, Casey JD, Wright A, Wang L, Shotwell MS, Siemann JK, et al; Vanderbilt Center for Learning Healthcare and the Pragmatic Critical Care Research Group. Cefepime vs Piperacillin-Tazobactam in Adults Hospitalized With Acute Infection: The ACORN Randomized Clinical Trial. JAMA 2023;330(16):1557–1567.

Introduction

Choosing empirical antibiotics requires clinicians to act before the infecting organism and its susceptibility are known. When antipseudomonal coverage is appropriate, cefepime and piperacillin–tazobactam are common options. Their coverage is not identical, and some infections clearly favour one regimen. However, when either is acceptable, comparative safety becomes an important part of the choice. Avoiding one possible drug toxicity should not inadvertently increase another.¹

Piperacillin–tazobactam had become associated with acute kidney injury (AKI), particularly when administered with vancomycin. Much of that evidence was observational. Antibiotic selection can track infection source, illness severity, renal function and other treatment decisions, making residual confounding difficult to exclude. There was also a mechanistic uncertainty: an increase in serum creatinine could reflect altered tubular secretion rather than an equivalent reduction in glomerular filtration. A prospective biomarker study had supported this possibility, but could not establish the comparative safety of the drugs by itself.^{1,2}

Cefepime raised a different concern. Neurological dysfunction, including delirium, reduced consciousness and seizures, had been described during treatment. In acutely ill patients these manifestations have many possible causes, including infection, organ failure, sedatives and other medicines. Drug concentration and kidney function may contribute, but the existence of recognisable toxicity cases does not quantify the average risk when cefepime is selected routinely. A comparative trial needed to assess neurological outcomes alongside kidney injury.^{1,3}

ACORN, Antibiotic Choice on Renal Outcomes, embedded that comparison within electronic prescribing at a large academic hospital. Randomisation occurred when clinicians were initiating empirical therapy and regarded both antibiotics as reasonable. The trial sought to determine whether cefepime reduced AKI or death compared with piperacillin–tazobactam, while also examining neurological dysfunction. Its practical

importance lies in the frequency of this initial prescribing decision, rather than in testing a new antimicrobial or prescribing the same drug for every suspected infection.¹

Synopsis

ACORN was an investigator-initiated, pragmatic, open-label, parallel-group randomised comparative safety trial at Vanderbilt University Medical Center in the United States. Adults in the emergency department or medical ICU were enrolled between 10 November 2021 and 7 October 2022; final follow-up was on 4 November 2022. An independent monitoring board oversaw the trial. Institutional and National Institutes of Health support funded the work, with no funder role in design, conduct, analysis, interpretation or publication decisions. The institutional review board approved a waiver of informed consent because both treatments were usual-care options and prospective consent could impede urgent antibiotic treatment.^{1,4}

Patients were eligible when a clinician initiated an order for cefepime or piperacillin–tazobactam within 12 hours of hospital presentation. Exclusions included penicillin or cephalosporin allergy, more than one dose of an antipseudomonal penicillin or cephalosporin during the previous seven days, imprisonment, and a clinical determination that one drug was required or preferable. An electronic tool screened patients and confirmed eligibility with the prescriber. Simple randomisation allocated treatment 1:1 without stratification, with assignment concealed until enrolment. Patients and clinicians were subsequently unblinded.¹

Electronic decision support recommended doses according to estimated kidney function. The standard cefepime regimen was 2 g intravenously over five minutes every eight hours. Piperacillin–tazobactam began with 3.375 g over 30 minutes, followed by 3.375 g every eight hours infused over four hours. Both schedules were adjusted for impaired kidney function. For example, cefepime frequency decreased to every 12 hours at an estimated filtration rate of 30–59 ml/min, while piperacillin–tazobactam maintenance dosing decreased to every 12 hours below 20 ml/min. Clinicians determined duration, additional antibiotics, escalation and discontinuation. Electronic reminders operated for seven days when assigned treatment was changed. Pharmacists assisted clinical dosing and monitored vancomycin levels; β -lactam concentrations were not routinely measured.^{1,4}

The primary endpoint was the highest stage of new or worsening AKI, or death, between randomisation and day 14, censored at hospital discharge. Its five ordered categories were survival without incident AKI, survival with stage 1, stage 2 or stage 3 AKI, and death. Kidney Disease: Improving Global Outcomes creatinine criteria were used,

including new kidney replacement therapy (KRT) for stage 3; urine output was not used. Pre-existing AKI alone did not count as a new event. Subsequent creatinine deterioration was required, and staging considered pre-illness, peri-enrolment and earlier on-study values. Patients already receiving KRT could enter only the survival or death categories.^{1,4}

There were two prespecified secondary outcomes. Major adverse kidney events at day 14 combined death, new KRT or a final inpatient creatinine at least twice baseline. Days alive and free of delirium and coma used positive Confusion Assessment Method for the ICU assessments and Richmond Agitation–Sedation Scale scores of –4 or –5. The supplement specifies a “last-off” calculation: free days counted from the final episode, with zero assigned after death or unresolved dysfunction at the endpoint. Patients discharged without dysfunction were assumed to remain free of it. Thus, this was not simply the sum of every normal neurological day. Delirium assessments were routine in ICU; sedation and consciousness scores were collected throughout the hospital.^{1,4}

The initial target was 2,050 patients, giving 80% power to detect a proportional odds ratio of 0.65 at a two-sided α of 0.05. After a planned interim review at 1,025 patients, the target increased to 2,500 to strengthen assessment among patients receiving vancomycin. Updated assumptions were 70% without AKI, 10%, 7% and 7% with successive AKI stages, and 6% dying. These gave 92% power for an odds ratio of 0.75, approximately a five-percentage-point difference in any AKI or death. The interim efficacy boundary was $P \leq 0.001$; the final threshold remained $P < 0.05$.^{1,4}

The primary population comprised randomised patients receiving at least one dose of either study antibiotic, analysed by assignment. An unadjusted proportional odds model compared the ordinal kidney endpoint; an odds ratio below one favoured cefepime. The same model family assessed neurological free days, where an odds ratio above one favoured cefepime. Adjusted and sensitivity analyses examined baseline severity, creatinine definitions and treatment exposure. Secondary and exploratory confidence intervals were not adjusted for multiplicity. A survivor discharged without any post-enrolment creatinine measurement was assumed not to have developed AKI; secondary outcomes used available complete cases.^{1,4}

Of 3,806 patients assessed, 2,634 were randomised. Four were subsequently excluded because they were incarcerated and 119 received neither study drug, leaving 1,214 cefepime and 1,297 piperacillin–tazobactam patients in the primary analysis. Median age was 58 years, 42.7% were female and 94.7% entered in the emergency department, at a median 1.2 hours after presentation. Sepsis was present in 54.2%. Median Sequential Organ Failure Assessment (SOFA) score was two in each group; approximately 13% were receiving vasopressors. Baseline renal disease was common, and 145 patients had

already received KRT. Mechanical ventilation, sedation and neurological dysfunction were somewhat more frequent in the cefepime group.^{1,4}

At least one assigned dose was administered to 95.1% and 98.4%, respectively. Median assigned-antibiotic duration was three days in both groups. Within 14 days, 18.8% assigned cefepime received piperacillin–tazobactam and 17.2% assigned piperacillin–tazobactam received cefepime. Vancomycin was being received at enrolment by 77.2% overall, and its median treatment duration was two days.¹

The primary ordinal comparison showed no significant difference: odds ratio 0.95 (95% CI 0.80–1.13; $P = 0.56$). In cefepime versus piperacillin–tazobactam groups, 910/1,214 versus 952/1,297 survived without incident AKI; 86 versus 100, 41 versus 70 and 85 versus 97 survived with stages 1, 2 and 3, respectively. Deaths by day 14 numbered 92 versus 78, 7.6% versus 6.0%. Adjustment and sensitivity analyses, including the population receiving no study drug, did not materially change the primary comparison.^{1,4}

Major adverse kidney events occurred in 124/1,214 versus 114/1,297 patients, 10.2% versus 8.8%, an absolute difference of 1.4 percentage points (95% CI –1.0 to 3.8). Mean neurological free days were 11.9 versus 12.2, with an odds ratio of 0.79 (95% CI 0.65–0.95). A post hoc analysis found delirium or coma in 252 versus 225 patients, 20.8% versus 17.3%, a difference of 3.4 percentage points (95% CI 0.3–6.6). Adjustment for baseline delirium or coma gave an odds ratio of 0.80 (95% CI 0.63–1.01) for free days. By day 28, censored at discharge, mortality was 8.6% versus 8.2%; new KRT occurred in 44/1,136 versus 28/1,230 eligible patients, an exploratory difference of 1.6 percentage points (95% CI 0.1–3.1). Allergic reactions were reported in 16 versus 15 patients.^{1,4}

Critique

ACORN challenges the assumption that selecting cefepime instead of piperacillin–tazobactam reliably prevents clinically important kidney harm during empirical treatment. Its randomised comparison is particularly valuable because the concern arose largely from non-randomised studies susceptible to treatment-selection bias. However, a non-significant superiority result does not prove renal equivalence or universal safety. The estimate applies to an initial treatment strategy with relatively short exposure, clinical adjustment and substantial switching, rather than to uninterrupted courses of either drug.¹

The ordinal primary outcome improves on a creatinine-only endpoint by ranking death as worst. It nevertheless compresses clinically different events into one summary odds ratio. Lower-stage AKI was numerically less frequent with cefepime, while death was

numerically more frequent. Post hoc analyses at successive thresholds gave odds ratios from 0.92 for any AKI or death to 1.28 for death alone, with all confidence intervals crossing one. The overall estimate is therefore not evidence that each kidney stage and mortality moved identically. The absence of a favourable signal in major adverse kidney events is an important complement to the primary result.^{1,4}

The outcome also answers a more specific question than the label “highest AKI stage” might imply. Patients with severe renal dysfunction at entry needed a new rise after enrolment to contribute an incident AKI event. Prior KRT excluded subsequent renal-stage transitions altogether. This is sensible when attributing new injury to a newly started antibiotic, but it means category zero does not necessarily describe normal kidneys. Creatinine changes after an initial improvement could qualify, whereas persistent pre-existing injury without further deterioration might not. The detailed definition is essential when comparing ACORN's event rates with observational cohorts using different baselines.⁴

Creatinine ascertainment was pragmatic rather than scheduled for research. A substantial minority lacked a measured pre-illness value, and an estimation equation supplied a substitute. Survivors discharged without a subsequent creatinine were assumed not to have AKI. These choices may miss mild or unmeasured injury. Nevertheless, analyses restricted to measured baseline creatinine, using alternative baseline definitions and excluding pre-existing AKI did not reveal a clear opposing result. Routine sampling patterns appeared similar between groups. These checks support the overall finding while leaving uncertainty about subtle kidney injury that would require other biomarkers or systematic follow-up.^{1,4}

The neurological result warrants attention, but its interpretation should be more measured than declaring every excess episode to be cefepime neurotoxicity. It was a prespecified secondary outcome without multiplicity adjustment, assessed by clinicians who knew treatment assignment. Delirium was systematically screened in ICU rather than across all wards. The trial did not collect the complete spectrum of cefepime-associated manifestations, such as myoclonus or seizures, or drug concentrations that might support individual causal attribution. It provides randomised evidence of a possible difference in clinically relevant neurological outcomes, with less certainty about mechanism.^{1,4}

Moreover, the free-day outcome incorporates death and the timing of the final abnormal assessment. Earlier normal days are not counted if dysfunction recurs. A small average difference across a population in which most patients score 14 days does not mean that each patient experienced a fraction of a day of drug toxicity. The accompanying excess in

patients with any delirium or coma supports a neurological contribution, but that component analysis was post hoc. The outcome can be affected by sedation, severity and discharge patterns as well as a direct antibiotic effect.^{1,4}

Baseline imbalance is relevant here. Cefepime patients more often received mechanical ventilation and continuous sedation and more often had coma or delirium around enrolment. Randomisation makes systematic allocation bias unlikely, but does not guarantee exact balance. Prespecified adjustment and additional sedation adjustment retained a similar point estimate. Adding baseline delirium or coma produced an interval extending just across one. This does not erase the original result, nor does it prove confounding explains it; it illustrates that the neurological estimate is modest and its statistical certainty sensitive to analysis choices.^{1,4}

Treatment separation was substantial initially but incomplete over time. Almost one fifth of each group received the alternative drug, and median assigned treatment lasted only three days. These features reduce confidence about the comparative toxicity of sustained exposure. They also reflect the clinical question: empirical antibiotics should change when new diagnostic information becomes available. Analyses restricted to longer courses did not show renal benefit with cefepime, but duration is determined after randomisation. Such restrictions select patients whose illness and treatment decisions permitted continued therapy and cannot substitute for randomising a prolonged course.^{1,4}

The primary analysis similarly excluded patients who never received either antibiotic. This modified population is reasonable for a drug-safety question, but treatment receipt occurs after assignment. The sensitivity analysis restoring those patients is therefore important, and its near-identical result reduces concern that the exclusion drove the finding. Four incarcerated patients remained excluded from data collection and analysis; the sensitivity population was not literally every randomised participant. Concealed electronic allocation and enrolment very early in care are concrete strengths, limiting clinician selection of assignment and extensive exposure before randomisation.^{1,4}

Several subgroup interactions reached nominal significance, including sepsis status and baseline neutropenia for the primary endpoint. With multiple candidate modifiers, small subgroups and no confirmatory trial, these do not establish prescribing rules. The vancomycin subgroup is particularly relevant to the original concern: its primary odds ratio was 0.89 (95% CI 0.73–1.08), without clear evidence of a renal advantage for cefepime. Vancomycin itself was not randomised, and this comparison does not determine the safety of longer combination treatment or different vancomycin exposure targets.⁴

Finally, most participants were initially treated outside ICU and many had low organ-failure scores. The trial is directly relevant to early empirical prescribing when both drugs are suitable, with greater uncertainty in prolonged critical illness. Five-minute cefepime administration, extended-infusion piperacillin–tazobactam and available pharmacy support are part of the tested setting. Antimicrobial efficacy, pathogen susceptibility and additional anaerobic coverage still matter. ACORN cannot justify selecting an inadequately active regimen on safety grounds, but it weakens kidney-toxicity concerns as an automatic reason to prefer cefepime and supplies a neurological reason for caution.^{1,4}

Body of Evidence

The observational evidence motivating concern about piperacillin–tazobactam was substantial but not uniform. Blevins and colleagues studied 2,492 ICU patients receiving vancomycin with piperacillin–tazobactam, cefepime or meropenem. AKI occurred in 39.3%, 24.2% and 23.5%, respectively; the adjusted association for the piperacillin–tazobactam combination had an odds ratio of 2.161 (95% CI 1.620–2.883). This was an important safety signal, but the groups differed in ICU location, sepsis, hypotension and vasopressor exposure. Adjustment could address recorded features without reproducing random allocation. The study made a comparative trial necessary; it could not establish that switching antibiotics would remove the observed excess risk.⁵

Miano and colleagues examined a different explanation in a prospective cohort of 739 critically ill patients receiving vancomycin with piperacillin–tazobactam or cefepime. Cystatin C was measured in 192. Piperacillin–tazobactam was associated with more creatinine-defined AKI, with a rate ratio of 1.34 (95% CI 1.01–1.78), but not with corresponding increases in cystatin C, dialysis or mortality. This supported the possibility that altered creatinine handling contributed to the association. The biomarker subset and observational allocation limit certainty, and the study does not prove that all creatinine rises are harmless. It explains why outcomes such as new KRT and persistent dysfunction deserve attention alongside transient biochemical changes.²

The neurological concern had its own observational basis. Boschung-Pasquier and colleagues examined patients undergoing cefepime therapeutic drug monitoring. Among 319 with available trough concentrations, presumed neurotoxicity occurred in 74, and concentrations were higher in those affected: median 21.6 versus 6.3 mg/l. Renal function was also poorer in affected patients. This supports a concentration-related toxicity hypothesis but does not provide a population incidence for routine empirical use. Selection for monitoring enriches the cohort for patients in whom accumulation or

toxicity is already a concern. ACORN adds randomised comparative evidence without establishing a concentration threshold or testing a monitoring intervention.³

A 2023 secondary analysis of SMART examined 3,199 critically ill adults who received piperacillin–tazobactam or an antipseudomonal cephalosporin within 24 hours of ICU admission. Initial piperacillin–tazobactam was not associated with more major adverse kidney events at 30 days: adjusted odds ratio 1.03 (95% CI 0.83–1.27). It was associated with more neurological free days, adjusted odds ratio 1.18 (95% CI 1.00–1.38), although that association did not remain statistically significant after additional medication adjustment. Crucially, SMART randomised fluid choice, not antibiotic choice; this antibiotic analysis remained observational. Its direction anticipated ACORN's results, but it was not an independent randomised confirmation of them.⁶

These studies address distinct parts of the safety question. Retrospective cohorts identify associations in selected treatment populations; biomarker studies test whether a laboratory signal is consistent with impaired filtration; concentration studies examine plausible drug toxicity; and pragmatic randomisation evaluates the consequences of choosing one regimen in clinical practice. Agreement across some of these observations strengthens a hypothesis, but their different populations and outcome definitions should not be collapsed into an apparently precise universal effect. ACORN's main advance is to remove much of the baseline treatment-selection problem at the moment antibiotics are started.^{1,2,3,5,6}

By the end of 2023, the resulting clinical position was more balanced than a categorical preference based on suspected nephrotoxicity. In patients for whom either regimen was appropriate, ACORN did not demonstrate that piperacillin–tazobactam caused worse kidney outcomes and suggested a modest neurological disadvantage with cefepime. It did not establish interchangeable efficacy, equivalence across all renal outcomes or safety during every duration of treatment. Choice should remain tied to the infection being treated, with renal dosing, review of continuing antibiotic need and attention to new neurological dysfunction. The evidence changes the weighting of safety concerns within that decision rather than eliminating the need for it.¹

Does choosing cefepime over piperacillin–tazobactam improve safety during early empirical treatment?

No renal advantage was demonstrated. Cefepime was associated with fewer days alive and free of delirium or coma.

References

1. Qian ET, Casey JD, Wright A, Wang L, Shotwell MS, Siemann JK, et al; Vanderbilt Center for Learning Healthcare and the Pragmatic Critical Care Research Group. Cefepime vs Piperacillin-Tazobactam in Adults Hospitalized With Acute Infection: The ACORN Randomized Clinical Trial. *JAMA* 2023;330(16):1557–1567.
2. Miano TA, Hennessy S, Yang W, Dunn TG, Weisman AR, Oniyide O, et al. Association of vancomycin plus piperacillin–tazobactam with early changes in creatinine versus cystatin C in critically ill adults: a prospective cohort study. *Intensive Care Med* 2022;48(9):1144–1155.
3. Boschung-Pasquier L, Atkinson A, Kastner LK, Banholzer S, Haschke M, Buetti N, et al. Cefepime neurotoxicity: thresholds and risk factors. A retrospective cohort study. *Clin Microbiol Infect* 2020;26(3):333–339.
4. Qian ET, Casey JD, Wright A, et al. Supplement 2 to: Cefepime vs Piperacillin-Tazobactam in Adults Hospitalized With Acute Infection: The ACORN Randomized Clinical Trial. *JAMA*. 2023.
5. Blevins AM, Lashinsky JN, McCammon C, Kollef M, Micek S, Juang P. Incidence of Acute Kidney Injury in Critically Ill Patients Receiving Vancomycin with Concomitant Piperacillin-Tazobactam, Cefepime, or Meropenem. *Antimicrob Agents Chemother* 2019;63(5):e02658-18.
6. Qian ET, Wang L, Stollings JL, Casey JD, Rice TW, Semler MW. Piperacillin-Tazobactam Versus Anti-Pseudomonal Cephalosporins and Renal and Neurologic Outcomes in Critically Ill Adults: A Secondary Analysis of the SMART Trial. *J Intensive Care Med* 2023;38(12):1127–1135.

MERCY

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.

Introduction

Choosing an antibiotic does not settle how best to administer it. Meropenem is frequently used in critically ill patients with suspected or established severe bacterial infection, yet the same daily dose can produce markedly different concentration profiles when delivered intermittently or continuously. Its time-dependent antibacterial activity provides a rationale for maintaining exposure above the pathogen's minimum inhibitory concentration (MIC). Critical illness complicates this objective through changes in drug distribution and clearance, while reduced bacterial susceptibility can narrow the margin between adequate and inadequate exposure. A continuous infusion might therefore improve treatment without changing the molecule or increasing its daily dose.¹

That rationale has practical qualifications. A loading dose is needed to establish exposure promptly; renal dysfunction and changing clearance still require attention; and continuous delivery depends on compatible intravenous access, preparation procedures and drug stability. Sustaining an insufficient concentration will not make an inactive antibiotic effective. Conversely, if intermittent treatment already provides adequate exposure, a smoother concentration profile may add little. The clinical question concerns the consequences of an administration strategy, rather than whether its pharmacokinetic rationale is plausible.^{1,2}

Earlier randomised studies offered encouraging pharmacological and clinical signals, but their findings were inconsistent. Trials differed in antibiotic selection, total dose, outcome definition and the populations retained for analysis. Improvements in concentration targets or microbiological clearance did not consistently translate into better survival or fewer days in intensive care. These distinctions matter particularly when a proposed treatment change could become routine across a heterogeneous population.^{3,4,5,6}

MERCY examined continuous versus intermittent meropenem administration in adults with sepsis requiring intensive care. Its distinctive hypothesis was that continuous administration could reduce a composite of death and the emergence of organisms with

extensive or pandrug resistance. The trial thus attempted to connect individual treatment effectiveness with the clinically consequential problem of antimicrobial resistance.¹

Synopsis

MERCY was a double-blind, parallel-group, randomised superiority trial conducted in 31 intensive care units (ICUs) at 26 hospitals in Croatia, Italy, Kazakhstan and Russia. Recruitment ran from 5 June 2018 to 9 August 2022, with final 90-day follow-up in November 2022. The Italian Medicines Agency funded the trial and approved its design, but had no role in its subsequent conduct, analysis or publication. Adults were eligible when the treating clinician decided to initiate meropenem for sepsis or septic shock. Sepsis required suspected or documented infection, systemic inflammatory response criteria and a Sequential Organ Failure Assessment (SOFA) score of at least two. Shock additionally required vasopressors to maintain mean arterial pressure of at least 65 mmHg and lactate above 2 mmol/L despite adequate fluid resuscitation.^{1,2}

Important exclusions included ongoing carbapenem treatment, relevant antibiotic allergy, a high Simplified Acute Physiology Score II (SAPS II; at least 65 in the principal report), AIDS, and prolonged immunosuppressive or corticosteroid treatment. Consent procedures followed local ethical requirements. A central web system allocated patients equally using computer-generated permuted blocks stratified by centre. Patients, treating physicians and investigators were blinded. Pharmacy and designated nursing personnel prepared treatment without participating in outcome collection or analysis. Each patient received both continuous and intermittent preparations, one containing meropenem and the other matching saline placebo.^{1,2}

Both groups received an initial 1 g loading dose. Standard maintenance treatment was 3 g over 24 hours by continuous infusion or 1 g every eight hours by intermittent infusion, described in the principal report as lasting 30–60 minutes. Maintenance dosing was reduced to 2 g daily with creatinine clearance below 50 mL/min; clinicians could double doses in selected circumstances, including high MICs or particular infection sites. The supplement specifies more frequent intermittent doses during the first 24 hours, a first-day detail not fully reconciled with the principal description. Continuous standard-dose treatment used freshly prepared 500 mg portions in 50 mL saline every four hours. It was not a single preparation left infusing for an entire day. Treatment duration and adjustments remained clinical decisions, supported by trial recommendations, with study assessment extending to day 28.^{1,2}

Baseline blood cultures and specimens from the suspected infection site were obtained, preferably before the loading dose where not recently collected. Subsequent microbiological assessment included additional cultures and surveillance specimens. Other antibiotics and routine sepsis care were permitted. Meropenem concentrations were not routinely measured, so dosing was not adjusted to a measured pharmacokinetic target.^{1,2}

The primary endpoint was all-cause mortality or emergence of pandrug-resistant or extensively drug-resistant bacteria by day 28. Pandrug resistance meant resistance to all available antimicrobial classes intrinsically active against the organism; extensive resistance meant resistance to all but one or two classes. Resistance assessment used microbiological specimens beyond blood cultures. Patients dying within 48 hours were excluded from the denominator for the separately reported resistance component, but remained in the primary composite through its mortality component. Follow-up established primary outcome status and 90-day survival for all participants.^{1,2}

Four secondary outcomes were specified: days alive and free of antibiotics, days alive and outside ICU through day 28, 90-day mortality, and a cumulative SOFA-based measure through day 28. Patients dying by day 28 received zero antibiotic-free and ICU-free days. The cumulative organ-failure outcome was omitted from the main analysis because substantial data were missing, especially after ICU discharge. The individual primary components, lengths of stay and ICU readmission were examined as exploratory outcomes. Secondary and exploratory comparisons were not adjusted for multiplicity.^{1,2}

A sample of 600 patients was intended to provide at least 80% power to detect a reduction in the primary event rate from 52% to 40%, using a two-sided significance level of 0.05. Three interim analyses were planned after 150, 300 and 450 participants, with conservative efficacy boundaries and conditional-power assessment for futility. The principal analysis retained randomised patients in their assigned groups and compared proportions using a χ^2 test, reporting an unadjusted relative risk and 95% confidence interval. Supporting analyses included adjusted logistic models, a per-protocol population and prespecified subgroup comparisons. The principal sample-size target was reached.^{1,2}

Of 896 patients assessed, 607 were randomised: 303 to continuous and 304 to intermittent administration. The main reasons for exclusion were existing carbapenem treatment and high SAPS II. Mean age was approximately 64 years, one-third were women, and 61% had septic shock. Median SOFA score was nine in both groups. Approximately three-quarters had a tracheal tube or tracheostomy. Randomisation occurred a median of five days after ICU admission in each group, indicating that this was

often a later hospital treatment decision. No pathogen was identified in 28% versus 30%. Among tested Gram-negative isolates, 35% versus 30% were not susceptible to carbapenems; these are isolate-level proportions, not percentages of randomised patients.¹

Median meropenem treatment duration was 11 days in both groups and the median daily dose was 3 g. Other antibiotics were recorded around enrolment in approximately three-quarters of participants with available data. Only one patient crossed from continuous to intermittent treatment, by error. Short interruptions occurred in three versus two patients, and allocation was unblinded in four versus two. Twelve patients subsequently found to meet exclusion criteria remained in the intention-to-treat analysis and were omitted from the reported per-protocol analysis.^{1,2}

The primary outcome occurred in 142/303 patients (46.9%) versus 149/304 (49.0%): absolute difference -2.1 percentage points (95% CI -10.1 to 5.8), relative risk 0.96 (95% CI 0.81–1.13), $P = 0.60$. Day-28 mortality was 91/303 (30%) versus 99/304 (33%), relative risk 0.92 (95% CI 0.73–1.17), $P = 0.50$. Newly detected extensive or pandrug resistance occurred in 68/288 (24%) versus 70/280 (25%), relative risk 0.94 (95% CI 0.71–1.26), $P = 0.70$.¹

By day 90, 127 patients had died in each group: relative risk 1.00 (95% CI 0.83–1.21), $P = 0.97$. Median antibiotic-free days were three versus two, with a reported mean difference of 0.4 days (95% CI -0.9 to 1.7), $P = 0.57$. Median ICU-free days were zero in both groups; mean difference 0.6 days (95% CI -1.0 to 2.2), $P = 0.40$. The per-protocol primary estimate was similar, relative risk 0.96 (95% CI 0.81–1.14). Prespecified subgroup analyses did not establish treatment-effect heterogeneity. No seizures or allergic reactions attributed to the study drug were reported.^{1,2}

Critique

MERCY did not demonstrate that continuous meropenem improved its composite primary outcome, but the result is less definitive than a claim that administration mode is irrelevant. Its confidence interval permits an absolute benefit of approximately ten percentage points and harm approaching six points. The trial reached its planned sample, yet that sample was designed around a large 12-point reduction. Failure to achieve the anticipated effect is informative; it does not establish equivalence or exclude a smaller worthwhile advantage. The 90-day mortality interval likewise includes clinically meaningful benefit and harm. The most defensible conclusion concerns the absence of demonstrated superiority for the tested strategy.¹

The comparison itself was unusually well protected from treatment-preference bias. Concealed allocation, matching continuous and intermittent preparations, very little crossover and complete survival follow-up support a credible causal estimate. Blinding was especially useful where clinician decisions could influence antibiotic duration, additional treatment, specimen collection or ICU discharge. Renal adjustments and discretionary dose increases did not erase the assigned mode of delivery, and median treatment continued for 11 days. This was a sustained contrast in administration, rather than a brief exposure followed by extensive switching.^{1,2}

However, sustained administration contrast is not proof of pharmacological separation. The trial did not measure whether continuous treatment improved the proportion of time above a relevant MIC, nor whether intermittent treatment was commonly inadequate. Both regimens could have achieved effective concentrations for susceptible organisms; neither could necessarily overcome organisms with resistance beyond achievable exposure. Without concentration data, the neutral clinical result cannot distinguish these possibilities from an absence of clinical benefit despite better target attainment. MERCY evaluates a practical administration policy, not individually optimised, concentration-guided therapy.¹

The resistance component adds importance and complexity. Preventing a resistant organism is desirable, but death and a positive culture do not have interchangeable consequences for patients. Moreover, new detection of an extensively resistant organism is not necessarily acquisition of resistance by the bacterium responsible for the original infection. It can reflect a different organism or detection through surveillance rather than a new clinically manifest infection. The supplementary microbiological accounts include anal swabs and cases without a consequent change in antibiotics. Thus the composite captures a broad mixture of fatal illness and microbiological events, rather than a single measure of meropenem treatment failure.^{1,2}

Ascertainment also depends on surviving and remaining available for microbiological sampling. Excluding deaths within 48 hours from the resistance component produced different denominators, 288 and 280, while all 607 patients remained in the composite analysis. Those early deaths were appropriately retained as primary events, but the separate resistance comparison is conditional on early survival. It should not be interpreted as an unrestricted randomised comparison of bacterial resistance in every enrolled patient. Complete determination of the composite and survival does not establish identical opportunities for culture detection after discharge. Blinding mitigates differential clinical investigation, but does not remove these structural features of the endpoint.^{1,2}

The population was clinically relevant but specific. Patients commonly entered several days into ICU admission, frequently after other antibiotic exposure, and some had no identified pathogen. This creates a demanding test in established critical illness, where subsequent survival may depend on processes no longer responsive to modest improvements in antibiotic exposure. It also broadens the practical question beyond culture-proven, susceptible Gram-negative infection. MERCY does not directly answer whether continuous treatment is beneficial from the first hours of community-onset sepsis, or in a selected population with measured underexposure and an organism whose MIC lies within an achievable therapeutic range.¹

Concomitant antibiotics further qualify the mechanism. Approximately three-quarters of patients were receiving another antibiotic around enrolment. Effective companion treatment could protect during periods of low meropenem concentration, reducing the incremental benefit of changing infusion mode. However, the listed drug classes do not establish that these agents were active against each patient's causative organism. It would be equally unjustified to assume universal microbiological protection or to dismiss the trial as a comparison overwhelmed by effective combination therapy. The relevant finding is that MERCY tested meropenem delivery within common, complex antimicrobial care, rather than guaranteed monotherapy.¹

Some reporting limitations affect reproducibility. The supplement's more frequent first-day intermittent schedule is not fully explained alongside the loading dose and eight-hourly regimen in the principal account. This uncertainty concerns early exposure, when prompt effective treatment may matter most, although subsequent maintenance doses and observed treatment duration are clearer. It should be clarified before using the publication as a detailed prescribing template. The trial also cannot quantify the ordinary resource difference between strategies: double-dummy blinding required every patient to receive both infusion formats. Operational feasibility was demonstrated within that supported arrangement, but routine comparative nursing workload or intravenous-access burden was not established.^{1,2}

The missing cumulative SOFA outcome is a substantive gap. Its absence cannot be summarised as a fourth negative secondary result. Missing observations after discharge are related to clinical course, while mortality also changes which organ-failure values can be observed. Supplementary first-week trajectories and selected-population cumulative analyses do not replace the planned full-population 28-day comparison. Conversely, this limitation does not invalidate the complete primary and mortality outcomes. It limits what can be concluded about the evolution of organ dysfunction and illustrates why apparently comprehensive endpoint lists require examination of the analyses actually delivered.^{1,2}

Supporting analyses add consistency without resolving the remaining uncertainty. Adjusted models and exclusion of the 12 ineligible participants left the primary conclusion unchanged. Subgroups based on high MICs, selected organisms, kidney injury and illness severity did not identify a convincing differential effect, but their smaller samples cannot establish equivalence within clinically attractive niches. Numerous additional exploratory analyses increase the chance of apparently favourable estimates. Finally, absence of reported drug-attributed seizures or allergy is reassuring within this trial; it is not a precise assessment of uncommon toxicity. The findings support a cautious judgement about efficacy, rather than a universal assertion of equal effectiveness and safety.^{1,2}

Body of Evidence

The preceding trials illuminate why a plausible pharmacological intervention remained unsettled. Chytra and colleagues randomised 240 critically ill patients to continuous or intermittent meropenem. Continuous treatment used a 2 g loading dose followed by 4 g daily, whereas intermittent treatment used 2 g every eight hours. Clinical success among evaluable patients was 83.0% versus 75.0% ($P = 0.180$); microbiological success was 90.6% versus 78.4% ($P = 0.020$). The clinical and microbiological analyses excluded patients who did not meet evaluability requirements, including some early deaths. The lower daily dose in the continuous group also means this was not simply an equal-dose comparison of administration mode. Its microbiological signal justified further investigation, but could not establish a patient-centred benefit for routine continuous treatment.³

The double-blind BLING pilot enrolled 60 patients across five ICUs, using piperacillin–tazobactam, meropenem or ticarcillin–clavulanate. In the pharmacokinetic analysis, concentrations exceeded the MIC in 18/22 patients receiving continuous treatment and 6/21 receiving intermittent treatment, 82% versus 29% ($P = 0.001$). Clinical cure was 70% versus 43% ($P = 0.037$), whereas ICU-free days and hospital survival did not differ significantly. The concentration analysis involved fewer patients than were randomised, and the small trial could not provide a reliable estimate of mortality benefit. Nevertheless, it offered direct evidence that changing delivery could alter antibiotic exposure, a mechanistic measurement absent from MERCY.⁴

BLING II then examined the strategy in 432 eligible participants across 25 ICUs. The primary outcome, days alive and outside ICU by day 28, was a median of 18 with continuous treatment and 20 with intermittent treatment ($P = 0.38$). Ninety-day survival was 74.3% versus 72.5%, and clinical cure was 52.4% versus 49.5%, without a statistically significant difference. Like the pilot, it allowed several β -lactam antibiotics. This broader

question differs from MERCY's molecule-specific comparison and resistance-containing primary endpoint, but its findings weakened the proposition that continuous infusion would routinely produce a large improvement in recovery from severe sepsis.⁵

BLISS provided a contrasting clinical signal in 140 patients at two centres, excluding those receiving renal replacement therapy. Clinical cure 14 days after antibiotic cessation was 56% with continuous treatment versus 34% with intermittent treatment ($P = 0.011$). Target attainment for concentrations above the MIC throughout the dosing interval was also higher: 97% versus 70% on day one and 97% versus 68% on day three. Short-term survival did not differ significantly. Because treatment was open-label and the principal endpoint was clinical cure, its favourable result answers a different question from a blinded comparison centred on mortality and resistant organisms. It supports pharmacological efficacy under its study conditions, without resolving the survival question.⁶

MERCY therefore adds a relatively large, blinded, meropenem-specific trial to an evidence base in which exposure, clinical cure and survival had not moved consistently together. It reduces confidence in a large average benefit from simply converting eligible ICU patients to continuous administration at broadly comparable maintenance doses. It does not refute the earlier pharmacokinetic observations, because it did not repeat those measurements. Nor does a neutral resistance-containing composite overturn a clinical-cure result as though the endpoints were identical. Their differences help explain the continuing uncertainty rather than justify selecting whichever study agrees with a preferred practice.¹

By the end of 2023, the evidence supported separating three decisions: whether the antibiotic is appropriate for the infection, whether its dose is adequate for the patient's clearance and pathogen, and how that dose should be delivered. MERCY addresses the third within a defined clinical strategy. Further clarification would require adequately sized patient-centred comparisons and better identification of patients in whom delivery mode changes effective exposure. Its contribution is to temper claims of routine clinical superiority while leaving a rational, testable role for more targeted investigation.

Does continuous meropenem improve outcomes in critically ill adults with sepsis?

MERCY did not demonstrate a reduction in death or newly detected extensive or pandrug resistance compared with intermittent administration. Smaller clinically important benefits remain uncertain.

References

1. 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.
2. MERCY Investigators. Supplement 3: MERCY Study Committees, eMethods, eTables, and eFigures. Supplement to Monti et al. *JAMA* 2023;330(2):141–151.
3. Chytra I, Stepan M, Benes J, Pelnar P, Zidkova A, Bergerova T, et al. Clinical and microbiological efficacy of continuous versus intermittent application of meropenem in critically ill patients: a randomized open-label controlled trial. *Crit Care* 2012;16(3):R113.
4. 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.
5. 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.
6. Abdul-Aziz MH, Sulaiman H, Mat-Nor MB, Rai V, Wong KK, Hasan MS, et al. Beta-Lactam Infusion in Severe Sepsis (BLISS): a prospective, two-centre, open-labelled randomised controlled trial of continuous versus intermittent beta-lactam infusion in critically ill patients with severe sepsis. *Intensive Care Med* 2016;42(10):1535–1545.
7. MERCY investigators. Trial protocol. Supplement 1 to: Continuous vs Intermittent Meropenem Administration in Critically Ill Patients With Sepsis. *JAMA*. 2023.
8. MERCY investigators. Statistical analysis plan. Supplement 2 to the MERCY randomised clinical trial; corrected version posted 13 October 2023. *JAMA*. 2023.

Possible Reporting Inconsistencies. The dosing descriptions differ. The main paper describes 1 g every eight hours and intermittent infusions over 30–60 minutes; the results appendix and posted protocol specify six-hourly dosing for the first 24 hours before eight-hourly dosing when creatinine clearance exceeds 50 mL/min, while the protocol and current corrected SAP define bolus infusion as less than 30 minutes. The first-day schedule and intermittent-infusion duration actually delivered are not clear from these accounts.^{1,2,7,8}

CLOVERS

National Heart, Lung, and Blood Institute Prevention and Early Treatment of Acute Lung Injury Clinical Trials Network; Shapiro NI, Douglas IS, Brower RG, Brown SM, Exline MC, Ginde AA, et al. Early Restrictive or Liberal Fluid Management for Sepsis-Induced Hypotension. *N Engl J Med* 2023;388(6):499–510.

Introduction

Persistent hypotension during early sepsis presents a familiar choice: give more intravenous fluid, start a vasopressor, or combine the two. Fluid may improve cardiac filling and flow, while vasopressors address loss of vascular tone. Both can cause harm. Additional fluid can accumulate in tissues and impair organ function; vasopressors may increase cardiac workload, provoke arrhythmia or compromise regional perfusion. The practical uncertainty concerns the balance and sequence of these treatments in a patient whose blood pressure remains low after initial resuscitation.¹

The 2021 Surviving Sepsis Campaign suggested at least 30 mL/kg of crystalloid within three hours for sepsis-induced hypoperfusion or septic shock, but classified this as a weak recommendation supported by low-quality evidence. It also encouraged dynamic assessment to guide subsequent fluid treatment. These recommendations left considerable room for judgement about when to stop giving fluid and prioritise vasopressors. Observational associations between larger fluid volumes and worse outcomes could not resolve that decision because sicker patients may both receive more fluid and be more likely to die.^{1,2}

Earlier vasopressor use offered a plausible means of limiting fluid exposure while restoring arterial pressure. Conversely, prioritising vasoconstriction before adequate intravascular filling could impair perfusion. By 2022, CLASSIC had found no survival advantage from fluid restriction in patients already receiving ICU care for septic shock. Its result did not settle the earlier emergency-department decision, when treatment trajectories and the need for intensive care were still evolving.^{1,3}

CLOVERS, the Crystalloid Liberal or Vasopressors Early Resuscitation in Sepsis trial, compared two explicit approaches during the first 24 hours after enrolment. One prioritised earlier vasopressors and less fluid; the other prioritised additional fluid, with vasopressors available as rescue. The hypothesis was that the restrictive strategy would reduce death before return home by day 90 in adults with sepsis-induced hypotension persisting after initial fluid treatment.¹

Synopsis

CLOVERS was an open-label, multicentre, randomised superiority trial at 60 US centres, funded by the National Heart, Lung, and Blood Institute within the Prevention and Early Treatment of Acute Lung Injury network. Recruitment occurred from 7 March 2018 to 31 January 2022. Adults required suspected or confirmed infection and hypotension after at least one litre of intravenous fluid. The supplementary eligibility criteria allowed systolic pressure below 100 mmHg, mean arterial pressure (MAP) below 65 mmHg, or ongoing vasopressor infusion, with blood pressure below the patient's usual baseline. Enrolment had to occur within four hours of meeting eligibility and within 24 hours of hospital presentation.^{1,4}

Patients were excluded after more than three litres of prior fluid, with pulmonary oedema or new fluid overload, with severe volume depletion from another acute condition, or when hypotension was attributed to a non-septic cause. Other exclusions included imminent limitations of most life-sustaining treatment, immediate surgery that prevented trial treatment, pregnancy and inability to obtain consent. Treating clinicians had to consider both strategies acceptable. Written consent came from patients or authorised representatives; recruitment depended on research-staff availability, usually during daytime and evening hours. A central web system randomised patients equally, stratified by site. Clinicians and patients were not blinded.^{1,4}

The restrictive strategy stopped maintenance and routine bolus fluids, while allowing discretionary boluses up to two litres in total, including fluid already received. For MAP below 65 mmHg or systolic pressure below 90 mmHg, norepinephrine was suggested and titrated to MAP at least 65 mmHg, with a second vasopressor if needed. Rescue 500 mL boluses were permitted for severe or refractory hypotension, rising lactate above 4 mmol/L after two hours, persistent marked tachycardia, or evidence of extreme hypovolaemia. Refractory hypotension included persistent low pressure despite norepinephrine at 20 µg/min or equivalent. Clinical judgement could override the algorithm.¹

The liberal strategy also stopped maintenance fluids but initially recommended a further two litres of isotonic crystalloid, normally completed within three hours. An October 2019 amendment introduced reassessment after the first litre: the second could be omitted if pressure and heart rate had normalised and the patient was judged volume replete. Subsequent 500 mL boluses responded to hypotension, increasing lactate, low urine output, tachycardia, vasopressor dependence or clinical and measured assessments of volume status. Rescue vasopressors were permitted for severe hypotension, rising lactate after two hours, fluid overload, or more than five litres of

total fluid. Fluids were stopped if overload developed. Clinicians could use vasopressors earlier when they considered this necessary.^{1,4}

Both approaches allowed exit from trial guidance for suspected central or peripheral ischaemia or mottling, and ended after 24 hours. The liberal algorithm also allowed return to care without trial guidance after at least five litres. Initial vasopressors could be delivered centrally or through a peripheral catheter of 20 gauge or larger. Trial personnel helped clinical teams implement the assigned strategy. The frequency of measurements and use of echocardiography or other haemodynamic assessments remained discretionary.¹

The primary outcome was death from any cause before discharge home by day 90. Home meant the same or a comparable setting to the patient's pre-illness residence; transfer from a private residence to rehabilitation did not count as return home. Patients discharged home or alive at day 90 were retained as non-events to day 91; those lost to follow-up were censored when last known alive. The primary analysis compared Kaplan–Meier day-90 point estimates using a z test, Greenwood standard errors and a 95% Wald confidence interval for the difference. It followed intention-to-treat allocation.¹

Secondary outcomes included organ-support-free days, days outside ICU and hospital, new organ support, kidney and organ-failure scores, and all-location 90-day mortality. The vasopressor-free-day analysis excluded the first 48 hours to allow for assigned treatment and a washout period. The reviewed results documents do not fully specify death scoring for every free-day measure. Safety assessment included new ventilation, arrhythmias, line complications and reported adverse events. Secondary and subgroup confidence intervals were not adjusted for multiplicity.^{1,4}

The planned sample was 2,320, providing 90% power to detect a reduction in the primary outcome from 15% to 10.5%, an absolute difference of 4.5 percentage points, at overall two-sided alpha 0.05. Interim reviews were planned around one-third and two-thirds of enrolment. Recruitment stopped after the second review on the monitoring board's recommendation for futility, with no identified safety concern. Its decision incorporated the primary and secondary outcomes, survival curves and further probability assessments; the appendix's numerical boundary table itself records an action of continuing at that review.^{1,4}

The flow diagram records 12,276 assessments and 4,868 patients meeting eligibility criteria. Ultimately, 1,563 were randomised, 782 to restrictive and 781 to liberal treatment. Consent difficulties, patient or surrogate refusal, and clinician refusal were

prominent reasons for non-enrolment. Mean age was 59.5 years, 47.2% were women, and 91.9% were enrolled in an emergency department. Mean trial SOFA score was 3.4, mean lactate among measured patients was 2.9 mmol/L, and approximately 7% were invasively ventilated. Chronic heart failure was present in 178 patients and dialysis-dependent kidney disease in 73. Median prior fluid volume was 2,050 mL in each group; 20.6% versus 18.3% were already receiving vasopressors.^{1,4}

Median post-randomisation fluid was 500 versus 2,300 mL during six hours, and 1,267 versus 3,400 mL during 24 hours. The estimated 24-hour difference in medians was -2,134 mL (95% CI -2,318 to -1,949). Including prior fluid, median cumulative volumes were 3,300 versus 5,400 mL. Vasopressors were administered during the first day to 460/780 patients (59.0%) versus 290/779 (37.2%). Among recipients with timing data, mean time from randomisation to first vasopressor was 1.8 versus 3.2 hours. Audited adherence was approximately 97% versus 96%. Subsequent daily fluid administration through day seven was similar.^{1,4}

Death before discharge home occurred in 109 versus 116 patients, yielding Kaplan–Meier estimates of 14.0% versus 14.9%: difference -0.9 percentage points (95% CI -4.4 to 2.6), $P = 0.61$. Five versus four patients were censored after loss to follow-up. All-location mortality by day 90 was 172/768 (22.4%) versus 169/773 (21.9%), difference 0.5 percentage points (95% CI -3.6 to 4.7). Mean organ-support-free days were 24.0 versus 23.6, with a reported difference of 0.3 days (95% CI -0.5 to 1.2). New invasive ventilation by day 28 occurred in 77/701 (11.0%) versus 87/687 (12.7%), and new renal replacement therapy in 24/738 (3.3%) in each group.¹

There were 21 versus 19 reported serious adverse events. New atrial or ventricular arrhythmias occurred in 59/779 (7.6%) versus 67/778 (8.6%). Three potential extravasations occurred among 500 patients receiving peripheral vasopressors; all resolved without intervention or residual consequences. A post hoc analysis found ICU admission during the first 24 hours in 525/780 (67.3%) versus 462/780 (59.2%), difference 8.1 percentage points (95% CI 3.3–12.8). Prespecified subgroup estimates did not identify a convincing subgroup-specific advantage.^{1,4}

Critique

CLOVERS addresses what to do after initial fluid administration, in patients for whom clinicians considered either further fluid or earlier vasopressors reasonable. This qualification defines the result. Both groups had received a median of approximately two litres before randomisation, and the restrictive group received a median total of 3.3 litres by the end of the intervention. The trial cannot establish that initial resuscitation

fluid is unnecessary, nor does it determine an optimal starting dose. Equally, it does not validate giving additional litres regardless of response: both strategies incorporated rescue treatment, reassessment and clinician override.¹

The achieved treatment contrast was substantial. A difference exceeding two litres during the first day, together with a 22-percentage-point difference in vasopressor exposure, shows that the interventions changed care. The neutral result is therefore difficult to dismiss simply as failure to separate treatments. However, fluid volume, vasopressor initiation and vasopressor duration changed together. This is a randomised comparison of two policies, not an isolated experiment on either fluid toxicity or the optimal timing of norepinephrine. Any causal effect belongs to the whole strategy, including its permitted exceptions.¹

The primary endpoint requires particular care. Death before returning home captures a clinically important failure of recovery and avoids treating transfer to rehabilitation as uncomplicated discharge. Yet once a patient returned home, subsequent death did not count in that primary endpoint. This is materially different from all-cause mortality through day 90. The distinction is visible in the results: primary estimates were approximately 14–15%, whereas all-location mortality was approximately 22%. Neither comparison demonstrated an advantage, but they should not be substituted for one another or combined into a single purported 90-day survival result.^{1,4}

Because return home depends partly on discharge decisions, rehabilitation access and pre-existing living arrangements, the primary measure combines survival with features of the recovery pathway. Open treatment could influence some of those decisions, although death itself is objective and only nine patients lacked complete primary follow-up. The separately reported all-location mortality provides useful reassurance that the primary result is not concealing an obvious survival difference after return home. Its wider interval and additional missing observations still limit precision. Mortality follow-up and discharge handling need to remain attached to the relevant endpoint.¹

Early termination reduced the information obtained. The trial stopped with about two-thirds of the planned sample, after the monitoring board judged that further recruitment was unlikely to establish the proposed benefit. The appendix makes clear that this was a broader judgement rather than automatic crossing of the tabulated futility boundary. Such oversight can reasonably consider more than one numerical rule, but futility for achieving a statistically significant result is not proof that the treatments are equivalent. The primary interval still permits a 4.4-point absolute benefit from

restriction or 2.6-point harm. These differences would matter clinically even though the anticipated 4.5-point benefit was not demonstrated.^{1,4}

Selection also shapes applicability. Patients with overt overload or severe non-septic volume depletion were excluded, as were those whose clinicians rejected either strategy. The eligible group was further filtered by consent and research availability. Randomisation protects the comparison among those enrolled; it does not make them representative of every patient with septic shock. Most entered from the emergency department, with relatively modest average organ dysfunction and no lactate threshold for inclusion. The result should therefore not be extended automatically to profound refractory shock, extreme fluid intolerance, delayed recognition or postoperative and hospital-acquired sepsis.^{1,4}

The liberal comparator was an active, structured approach. Its initial additional two litres and subsequent bolus triggers may differ from a clinician's usual response after reassessing the first litre. The 2019 amendment acknowledged that some patients became stable and volume replete before the second litre. After the amendment, this option was used in 38/282 liberal-group patients. Treatment separation persisted, and no convincing interaction by amendment period was shown. Nevertheless, CLOVERS compares the two specified algorithms; it cannot determine whether either improves on careful, unstructured clinical care or on a strategy consistently guided by measured fluid responsiveness.^{1,4}

Clinician discretion is both necessary protection and part of the estimand. Severe hypotension, overload or suspected ischaemia could prompt departure from the assigned pathway. These safeguards make the results relevant to monitored practice with rescue available, rather than rigid enforcement of a volume ceiling or vasopressor-first rule. Conversely, rescue and convergence after 24 hours may reduce differences in later outcomes. Similar fluid administration after the intervention does not invalidate the early comparison, but it limits conclusions about a restrictive policy maintained throughout ICU admission or a later strategy of active fluid removal.¹

The resource consequences deserve attention. More patients assigned to restriction entered ICU and underwent central line insertion. Greater vasopressor use could plausibly contribute to those differences, but the trial did not randomise ICU admission or formally establish that mechanism. These exploratory findings matter because a strategy without demonstrated survival benefit may still affect bed use, monitoring and procedures. The favourable experience with peripheral vasopressors is useful operational evidence, although catheter route was not randomised. Three uncomplicated extravasations among 500 selected recipients cannot establish equal

safety across catheter sites, durations, concentrations or less closely monitored settings.^{1,4}

Safety and subgroup findings do not overturn the principal uncertainty. Fluid overload and pulmonary oedema were reported more often as serious adverse events with liberal treatment, but unblinded attribution may influence reporting. Systematically collected ventilation and arrhythmia outcomes showed no clear difference, and rare ischaemic events remain imprecisely assessed. The small dialysis subgroup had a large point estimate favouring restriction with a wide interval crossing zero. A post hoc assisted-ventilation subgroup pointed in the opposite direction from the corresponding observation in CLASSIC. These are reasons for further study, not reliable bedside rules. CLOVERS supports flexibility between monitored strategies after initial resuscitation while leaving clinically important effects and patient selection unresolved.^{1,4}

Body of Evidence

Evidence from other settings first demonstrated why additional fluid could not be assumed harmless. In the *Zambian Simplified Severe Sepsis Protocol 2* trial, 209 adults were analysed after three post-randomisation exclusions. A six-hour strategy incorporating fluid boluses, vasopressors and transfusion increased in-hospital mortality from 33.0% to 48.1%, relative risk 1.46 (95% CI 1.04–2.05). Most participants had HIV, and virtually all were treated on wards without mechanical ventilation. This was a different population, care environment and multicomponent intervention from CLOVERS. It establishes a context-specific hazard, while warning against interpreting the US trial as evidence that liberal resuscitation is universally safe.⁵

CENSER tested early norepinephrine more directly in 310 adults with sepsis and hypotension at one Thai hospital. In this blinded phase II trial, shock control at six hours occurred in 76.1% with early norepinephrine versus 48.4% with standard treatment ($P < 0.001$). Twenty-eight-day mortality was 15.5% versus 21.9% ($P = 0.15$). The physiological endpoint supported the feasibility and haemodynamic effect of earlier treatment, but did not establish a survival advantage. CLOVERS extended the investigation to a larger multicentre policy comparison, changing fluid and vasopressor use together. Its neutral outcome tempers an inference from faster shock control to improved survival, without negating the earlier physiological finding.⁶

FRESH examined another approach: using the stroke-volume response to passive leg raising to guide fluid and vasopressor decisions. Its modified intention-to-treat analysis included 124 eligible patients from 150 randomised. Mean fluid balance at 72 hours or ICU discharge was 0.65 versus 2.02 litres, a difference of -1.37 litres ($P = 0.021$), with

favourable secondary signals for renal replacement therapy and ventilation. The small sample and selected primary analysis limit certainty about clinical outcomes. Its relevance is conceptual: a consistently applied response-guided strategy is a different intervention from choosing a generally restrictive or liberal algorithm. CLOVERS did not require dynamic testing before each treatment decision and therefore does not resolve FRESH's specific question.⁷

CLASSIC, published during the final year of CLOVERS follow-up, randomised 1,554 ICU patients with septic shock after at least one litre of fluid. Restricted treatment reduced median ICU intravenous fluid from 3,811 to 1,798 mL, but 90-day mortality was 42.3% versus 42.1%, adjusted absolute difference 0.1 percentage points (95% CI -4.7 to 4.9), $P = 0.96$. Its population had established ICU shock and substantially higher mortality than CLOVERS. Its primary endpoint counted all deaths through day 90, and its standard-care comparator differed from CLOVERS's liberal algorithm. The two trials consequently provide complementary evidence at different treatment stages, rather than exact replication of a single strategy.³

Taken together, these studies separate several questions that are often compressed into "more or less fluid". Earlier vasopressors can change haemodynamic recovery; dynamic assessment can reduce fluid accumulation; fluid-heavy care can be harmful in some settings; and substantial fluid restriction has not consistently improved survival in large trials conducted where intensive care support is available. These findings are compatible because the populations, co-interventions, timing and endpoints differ. They do not justify a universal volume prescription or the assumption that a particular physiological response is a validated surrogate for survival.^{1,3,5,6,7}

At the end of 2023, CLOVERS narrowed the uncertainty around the average effect of two early resuscitation policies after initial fluid treatment. It supported the feasibility of using less fluid and earlier vasopressors with appropriate monitoring, without demonstrating superiority for mortality or organ-support outcomes. The remaining task was to identify when additional fluid is likely to improve effective perfusion and when vasopressors should take priority. The trial leaves that individual assessment relevant; it does not replace it with a requirement to complete either algorithm regardless of the patient's response.

Should early vasopressors and fluid restriction routinely replace a liberal fluid strategy in sepsis-induced hypotension?

After initial fluid resuscitation, CLOVERS did not demonstrate better survival with the restrictive strategy. Either tested approach requires reassessment and access to rescue treatment; equivalence was not established.

References

1. National Heart, Lung, and Blood Institute Prevention and Early Treatment of Acute Lung Injury Clinical Trials Network; Shapiro NI, Douglas IS, Brower RG, Brown SM, Exline MC, Ginde AA, et al. Early Restrictive or Liberal Fluid Management for Sepsis-Induced Hypotension. *N Engl J Med* 2023;388(6):499–510.
2. 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.
3. Meyhoff TS, Hjortrup PB, Wetterslev J, Sivapalan P, Laake JH, Cronhjort M, et al; CLASSIC Trial Group. Restriction of Intravenous Fluid in ICU Patients with Septic Shock. *N Engl J Med* 2022;386(26):2459–2470.
4. CLOVERS Investigators. Supplementary Appendix to: Early Restrictive or Liberal Fluid Management for Sepsis-Induced Hypotension. *N Engl J Med* 2023;388(6):499–510.
5. Andrews B, Semler MW, Muchemwa L, Kelly P, Lakhi S, Heimbürger DC, et al. Effect of an Early Resuscitation Protocol on In-hospital Mortality Among Adults With Sepsis and Hypotension: A Randomized Clinical Trial. *JAMA* 2017;318(13):1233–1240.
6. Permpikul C, Tongyoo S, Viarasilpa T, Trainarongsakul T, Chakorn T, Udompanturak S. Early Use of Norepinephrine in Septic Shock Resuscitation (CENSER). A Randomized Trial. *Am J Respir Crit Care Med* 2019;199(9):1097–1105.
7. Douglas IS, Alapat PM, Corl KA, Exline MC, Forni LG, Holder AL, et al. Fluid Response Evaluation in Sepsis Hypotension and Shock: A Randomized Clinical Trial. *Chest* 2020;158(4):1431–1445.

ACTIV-1 IM

O'Halloran JA, Ko ER, Anstrom KJ, Kedar E, McCarthy MW, Panettieri RA Jr, et al; ACTIV-1 IM Study Group Members. Abatacept, Cenicriviroc, or Infliximab for Treatment of Adults Hospitalized With COVID-19 Pneumonia: A Randomized Clinical Trial. JAMA 2023;330(4):328–339.

Introduction

The recognition that modifying the host response could improve survival changed the treatment of severe COVID-19. Dexamethasone established that anti-inflammatory treatment could benefit patients requiring respiratory support, while also demonstrating that the consequences of immunosuppression depended on the population treated. The subsequent question was whether more selective interventions could add benefit to corticosteroids without compromising defence against SARS-CoV-2 or secondary infection.¹

Several inflammatory pathways offered plausible targets, but plausibility did not establish which should be inhibited, when treatment should begin or which patients should receive it. Abatacept modifies T-cell costimulation; infliximab neutralises tumour necrosis factor α ; and cenicriviroc antagonises the chemokine receptors CCR2 and CCR5, affecting monocyte and macrophage function. These interventions therefore represented different biological hypotheses. Experience with abatacept and infliximab in other inflammatory diseases supported investigation, but could not establish their effectiveness or safety during acute viral pneumonia.²

The pandemic also demanded an efficient way to assess several treatments while usual care and the circulating virus were changing. A common trial framework with concurrent placebo controls could reduce duplicated recruitment and allow unsuccessful agents to stop early. Its interpretation would nevertheless require care: sharing controls does not make separate drug comparisons independent replications, and improvement in recovery is a different question from prevention of death.

ACTIV-1 Immune Modulator, part of the Accelerating COVID-19 Therapeutic Interventions and Vaccines programme, tested whether adding abatacept, cenicriviroc or infliximab to standard treatment accelerated recovery in adults hospitalised with COVID-19 pneumonia. Recruitment began in October 2020; evidence supporting other additional immunomodulators developed during the trial. The relevant clinical question

by publication was therefore both whether these agents worked and how convincingly they could enlarge the available treatment options.²

Synopsis

ACTIV-1 IM comprised three randomised, double-masked, placebo-controlled superiority comparisons under a common trial framework. Between 16 October 2020 and 31 December 2021, 1971 participants were randomised at 95 hospitals associated with 85 research sites in the United States, Peru, Argentina, Brazil and Mexico. Federal support came through the Biomedical Advanced Research and Development Authority and National Center for Advancing Translational Sciences. Manufacturers donated study drugs and remdesivir; public-sector investigators and sponsors contributed to trial conduct, and industry representatives participated in development discussions. An independent data and safety monitoring board supervised the trial.²

Adults required confirmed SARS-CoV-2 infection within 14 days, an anticipated hospital stay of at least 72 hours and pulmonary involvement. The latter could be demonstrated radiologically, by an oxygen saturation of 94% or less on room air, or by a requirement for oxygen or ventilatory support. Patients ranged from those still requiring inpatient care without oxygen to those receiving invasive ventilation or extracorporeal membrane oxygenation. Written consent could be provided by the participant or a legally authorised representative.^{2,3}

Important exclusions included pregnancy, major hepatic abnormalities, acute kidney injury with estimated glomerular filtration rate below 30 mL/min, severe heart failure, marked neutropenia or lymphopenia, untreated infection including tuberculosis, and recent targeted immunomodulation. Stable chronic renal insufficiency below that filtration threshold was permitted. The supplement explains that the liver-enzyme exclusion changed from more than five to more than ten times the upper limit of normal. Abatacept had no additional substudy-specific exclusions. Cenicriviroc excluded patients taking contraindicated medicines that could not be stopped or replaced; infliximab additionally excluded specified lymphoma histories and significant demyelinating disease.³

Allocation occurred in two stages. First, participants were openly assigned with equal probability to one of the available agents for which they were eligible. They were then randomly assigned, under masking, to that agent or its matching placebo in an n:1 ratio, where n was the number of agents for which they were eligible. Randomisation was stratified by geographical location and moderate or severe disease. Placebo recipients contributed to an agent's comparison only if they met its eligibility criteria and enrolled

at a site and time when that agent was available. Thus, the three control groups overlapped but were not identical. The report describes double masking of the active-versus-placebo comparisons without providing every operational detail of sequence generation or allocation access.^{2,3}

Abatacept was administered intravenously on day 1 at 10 mg/kg, to a maximum of 1000 mg, against a matching saline infusion. Infliximab was given as a single 5 mg/kg infusion over at least two hours, also against saline. Cenicriviroc treatment usually began with 300 mg in the morning and 150 mg that evening, followed by 150 mg approximately every 12 hours with food for 28 days. Matching oral placebo was used; enteral-tube administration was permitted, and treatment continued after discharge if necessary.³

All groups received local supportive care. Remdesivir was supplied by the study and initiated in eligible patients not already receiving it, using a 200 mg intravenous loading dose followed by 100 mg daily for up to ten days. Corticosteroids were permitted in accordance with national guidance. Other immunomodulators were prohibited at randomisation, although subsequent tocilizumab and baricitinib use occurred. Clinical status was assessed during hospitalisation and at scheduled follow-up contacts; the final safety assessment was at day 60.^{2,3}

The primary outcome was time from randomisation to first recovery by day 28. Recovery meant attaining category 6, 7 or 8 on an eight-category scale: remaining in hospital without oxygen or ongoing medical-care requirements; being discharged with activity limitations or home oxygen; or being discharged without activity limitations. The lower categories distinguished death, invasive ventilation or extracorporeal support, non-invasive ventilation or high-flow oxygen, conventional oxygen, and continued inpatient care without oxygen. Higher values represented better status. The key secondary endpoints were day-14 clinical status and all-cause mortality through day 28. Additional outcomes included mortality at days 14 and 60 and clinical status at day 28.²

Each comparison required an estimated 788 recoveries for approximately 85% power to detect a recovery rate ratio of 1.25. The primary intention-to-treat analysis used a Fine–Gray model with death as a competing event, stratified by geographical region and baseline ordinal severity, with age and sex as covariates. The reported recovery rate ratio describes the modelled recovery process; values above one favour active treatment. Efficacy monitoring used Lan–DeMets spending with an O’Brien–Fleming-type boundary, and futility monitoring used Hwang–Shih–DeCani spending. Three interim analyses were planned, but only two occurred because recruitment was rapid.^{2,3}

A gatekeeping procedure required success on the primary endpoint before either key secondary endpoint could be declared statistically significant. Mortality was analysed by adjusted logistic regression, including region, baseline ordinal severity, age and sex; the two least severe entry categories were combined for model convergence. Clinical status used proportional-odds regression. Multiple imputation addressed missing outcomes, although the detailed imputation rules are outside the material reviewed here. The principal report presents secondary and safety analyses in participants who received at least one assigned dose, termed the treated-as-assigned population.^{2,3}

Of 2055 consenting participants, 1971 were randomised: 524 to abatacept, 360 to cenicriviroc, 531 to infliximab and 556 to placebo. Shared-placebo eligibility produced comparison groups of 525, 363 and 530, respectively. The population had a mean age of approximately 55 years; 61.8% were men. Around half required conventional oxygen, one-third high-flow oxygen or non-invasive ventilation, and roughly one-tenth invasive ventilation or extracorporeal support. Symptoms had generally begun nine to ten days earlier. Obesity was common, and more men entered active-treatment groups than their respective controls.²

The treated-as-assigned populations were 509 versus 510 for abatacept, 355 versus 354 for cenicriviroc, and 517 versus 516 for infliximab. Remdesivir use during days 1–5 was approximately 93–96%, and corticosteroid use approximately 89–93% in the randomised groups. Subsequent tocilizumab and baricitinib use was uncommon. In the cenicriviroc comparison, 98.4% received a loading dose and 96.7% at least one maintenance dose; median treatment duration was 28 days, although 36.5% discontinued maintenance before day 29. On 2 September 2021, the monitoring board stopped cenicriviroc randomisation for futility, while maintaining treatment, follow-up and masking for those already enrolled. Abatacept and infliximab recruitment ended after the required recovery totals were reached.²

None of the three comparisons met the primary endpoint. With abatacept, 414/524 participants recovered by day 28 versus 397/525 controls, 79.0% versus 75.6%; the recovery rate ratio was 1.12 (95% CI 0.98–1.28; $P=0.09$), with median recovery of nine days in both groups. For cenicriviroc, recovery occurred in 273/360 versus 287/363, 75.8% versus 79.1%; the rate ratio was 1.01 (95% CI 0.86–1.18; $P=0.94$), with medians of eight days in both groups. For infliximab, the corresponding figures were 421/531 versus 405/530, 79.3% versus 76.4%; the rate ratio was 1.12 (95% CI 0.99–1.28; $P=0.08$), with medians of eight versus nine days.²

In the treated-as-assigned analyses, 28-day mortality was 56/509 versus 77/510 with abatacept, 11.0% versus 15.1%, with an adjusted, imputed odds ratio of 0.62 (95% CI

0.41–0.94). For cenicriviroc, it was 49/355 versus 42/354, 13.8% versus 11.9%, odds ratio 1.18 (95% CI 0.72–1.94). For infliximab, it was 52/517 versus 75/516, 10.1% versus 14.5%, odds ratio 0.59 (95% CI 0.39–0.90). The odds ratios for better day-14 clinical status were, in the same order, 1.19 (95% CI 0.94–1.50), 0.93 (0.71–1.23) and 1.32 (1.05–1.66). These secondary intervals were not confirmatory results under the testing hierarchy.²

At day 60, mortality was 14.5% versus 17.1% with abatacept, 18.0% versus 13.8% with cenicriviroc, and 12.6% versus 16.5% with infliximab. The composite of death, serious adverse events or grade 3/4 adverse events occurred in 33.2% versus 34.9%, 36.9% versus 33.9%, and 32.9% versus 33.7%, respectively. All three confidence intervals for the absolute safety differences included no difference. Secondary infections through day 60 affected 16.1% versus 14.3%, 14.9% versus 14.7%, and 15.3% versus 14.0%. Three serious abatacept infusion reactions included one episode meeting criteria for anaphylaxis; one mild infliximab infusion reaction was reported.²

Critique

The main interpretative challenge is the discordance between unsuccessful primary testing and potentially important secondary mortality findings. ACTIV-1 IM did not demonstrate accelerated recovery with any agent. It nevertheless generated credible hypotheses that abatacept and infliximab might reduce mortality. Describing those drugs as proven to improve survival would exceed the analysis framework; dismissing their mortality findings entirely would also discard clinically relevant information. Their proper status is promising evidence requiring confirmation.²

The hierarchy was consequential, rather than a technical footnote. Selecting a favourable endpoint after the primary comparison fails increases the opportunity for chance findings across several outcomes and agents. Moreover, the supplementary gatekeeping tables give final recovery P values of 0.0574 for abatacept and 0.0631 for infliximab, corresponding to treated-as-assigned results, while the main primary analyses are intention-to-treat. Neither set crosses the reported efficacy boundary of $P < 0.0469$. This reporting distinction does not change the conclusion, but the estimates should not be silently interchanged. The secondary confidence intervals likewise should not be presented as if they incorporated successful hierarchical testing.³

Precision also matters. The primary intervals for abatacept and infliximab extend to a recovery rate ratio of 1.28, leaving the target effect of 1.25 compatible with the data. Thus, these are not equivalence results demonstrating that meaningful benefit is absent. Cenicriviroc provides a different pattern: its recovery estimate is near unity, the upper confidence limit is below the target effect, and its mortality estimates do not favour

treatment. Early futility stopping limits the information accumulated, but there is little positive evidence supporting this regimen. It does not follow that every possible manipulation of CCR2/CCR5 is ineffective.

The recovery endpoint captures an important aspect of illness but has limitations for this question. Patients discharged with home oxygen or substantial activity limitations were counted as recovered. A first qualifying recovery was also an event even if later deterioration occurred. Consequently, this endpoint measures a transition out of acute inpatient dependency, not complete or sustained restoration of health. Using death as a competing event was appropriate and avoided treating death as ordinary non-informative censoring. Even so, a model of recovery cannot substitute for a trial designed primarily to establish survival benefit.

The shared-control design was an important efficiency gain. Restricting controls by concurrent enrolment, site availability and drug eligibility protected comparisons against changes in care and differences in contraindications. Masked active-versus-placebo assignment further reduced the risk that knowledge of treatment would affect assessment or discharge decisions. These protections are especially relevant when recovery depends partly on clinical judgements. The open first allocation stage disclosed the substudy, however, and the available report does not establish masking of every operational role. Its design should be described accurately without assuming universal concealment throughout all processes.

Sharing controls also qualifies apparent replication. The abatacept and infliximab control populations overlapped substantially, so favourable findings in the two comparisons are statistically related. A common control-group experience can influence both. Conversely, cenicriviroc's unsuccessful comparison is not a validation test proving that the other agents work. Drug-specific eligibility, recruitment periods and treatment mechanisms differ. Comparing the mortality odds ratios across arms cannot establish which active agent is preferable; no reported head-to-head analysis answers that question.

The change of analysis population between primary and secondary outcomes deserves attention. Excluding participants who never received an assigned dose can weaken the protection conferred by randomisation if non-receipt is related to prognosis. The numbers excluded were modest, and supplementary intention-to-treat mortality analyses remained directionally similar. For example, infliximab's intention-to-treat 28-day mortality analysis included 53 deaths rather than the 52 in the treated-as-assigned analysis, with an odds ratio of 0.61 (95% CI 0.40–0.94). This is useful reassurance about sensitivity to population definition, while leaving multiplicity unresolved.³

Follow-up and missing-data assumptions remain relevant to the size of the mortality signal. The flow diagram records losses to follow-up and withdrawals in each comparison, and the numbers with observed ordinal assessments are smaller than the treated populations. Multiple imputation addresses missingness through assumptions; it does not turn uncertain outcomes into observed events. Because the protocol and statistical analysis plan were excluded from this appraisal, the detailed imputation model and sensitivity to alternative missing-data mechanisms could not be independently assessed. The chapter therefore retains the reported adjusted, imputed estimates without treating their robustness as fully established.^{2,3}

Treatment delivery supports an interpretable comparison, particularly for the two single infusions, for which prolonged outpatient adherence was unnecessary. Cenicriviroc required continued oral or enteral administration, including after discharge, and more than one-third stopped maintenance before day 29. This could dilute a treatment effect, but the published aggregate does not identify how much discontinuation represented modifiable non-adherence rather than illness or other circumstances. It cannot rescue an ineffective observed strategy through an assumed biological benefit among adherent patients.

Applicability is strongest to the hospital population actually recruited. Most patients were receiving oxygen without invasive ventilation, were around the second week of symptoms, and received both corticosteroids and remdesivir. Few received tocilizumab or baricitinib after randomisation. The trial therefore supplies little direct information about combining abatacept or infliximab with those agents, or replacing an established additional immunomodulator. Recruitment largely preceded the Omicron era, and vaccination status was unavailable. Absolute benefit in populations with different immunity and baseline mortality cannot be read directly from these event rates.²

Nor do the subgroup findings justify routine selection by age, diabetes or respiratory-support category. Numerous comparisons generate opportunities for apparently distinctive responses. Patients already receiving invasive ventilation or extracorporeal support were relatively few, and their mortality estimates did not favour abatacept or infliximab. The larger favourable differences among patients on high-flow oxygen or non-invasive ventilation are clinically interesting but do not establish a treatment rule. Statements that mortality benefit was uniform across severity groups would obscure these data.^{2,3}

Finally, similar aggregate safety results are reassuring within their limits. Infection ascertainment included independent review, and serious infusion reactions were explicitly reported. However, the trial excluded important high-risk groups, was not

designed to prove safety equivalence, and followed patients for only 60 days. Numerically higher infection frequencies in active-treatment groups, especially alongside lower mortality, require more careful interpretation than a declaration of no infection risk. The net balance of benefit and harm remains tied to the same uncertainty affecting efficacy.

Body of Evidence

The corticosteroid evidence established the foundation on which ACTIV-1 IM was built. In RECOVERY, 2104 patients received dexamethasone 6 mg daily for up to ten days and 4321 received usual care. Mortality at 28 days was 22.9% versus 25.7%, with an age-adjusted rate ratio of 0.83 (95% CI 0.75–0.93). Benefit was observed among patients receiving oxygen or invasive ventilation, but was not demonstrated among those receiving no respiratory support. This supported host-directed treatment while cautioning against extending it indiscriminately across disease severities. It also meant that a later trial needed to demonstrate benefit additional to effective background care.¹

Tocilizumab subsequently provided direct evidence for a more selective additional intervention. RECOVERY randomised 4116 patients with hypoxia and C-reactive protein of at least 75 mg/L; 82% were receiving systemic corticosteroids. Death by day 28 occurred in 621/2022 patients allocated tocilizumab and 729/2094 allocated usual care, approximately 31% versus 35%, with a rate ratio of 0.85 (95% CI 0.76–0.94). Mortality was the primary endpoint. Compared with ACTIV-1 IM, this was a larger, open-label comparison in a population selected for systemic inflammation and with higher mortality. These differences help explain why the apparent sizes of treatment effects cannot establish comparative effectiveness.⁴

COV-BARRIER illustrates why a secondary mortality finding can warrant further investigation after unsuccessful primary testing. In this double-blind trial of 1525 hospitalised adults, baricitinib did not significantly reduce the primary composite of progression to high-flow oxygen, non-invasive or invasive ventilation, or death: 27.8% versus 30.5%, odds ratio 0.85 (95% CI 0.67–1.08). However, 28-day mortality was 62/764 versus 100/761, with a hazard ratio of 0.57 (95% CI 0.41–0.78) and a nominal P value of 0.0018. Most participants were receiving corticosteroids. The parallel with ACTIV-1 IM is the need to preserve the original endpoint hierarchy while recognising that a potentially important survival effect deserves a dedicated evidential response.⁵

That response for baricitinib included a much larger RECOVERY comparison. Among 8156 participants, 95% were receiving corticosteroids and 23% tocilizumab at

randomisation. Mortality, again the primary endpoint, occurred in 514/4148 patients allocated baricitinib and 546/4008 allocated usual care, with an age-adjusted rate ratio of 0.87 (95% CI 0.77–0.99). This supported a mortality benefit whose estimated proportional magnitude was smaller than in earlier trials. For ACTIV-1 IM, the lesson is methodological as well as clinical: subsequent randomised confirmation can strengthen a promising signal and refine its size. Evidence that baricitinib works does not itself confirm abatacept or infliximab.⁶

By the end of 2023, therefore, the decision was no longer simply whether to add any anti-inflammatory drug to supportive care. Corticosteroids and additional targeted immunomodulation already had substantial randomised support. ACTIV-1 IM investigated different pathways and could potentially broaden options where contraindications or availability constrained treatment. Its placebo comparisons, however, did not establish equivalence to tocilizumab or baricitinib, effectiveness in combination with them, or a reliable hierarchy among the three experimental agents.

The most useful next evidence would distinguish confirmation of survival benefit from refinement of treatment selection. A mortality-focused randomised comparison of abatacept or infliximab, against care reflecting established additional immunomodulation, would address the principal uncertainty. Adequate representation of invasively ventilated patients, documentation of vaccination and viral era, and assessment of serious infection would clarify applicability. Biomarker work could help generate selection hypotheses, but a subgroup signal alone should not determine treatment. Within the book's evidence horizon, ACTIV-1 IM advances the case for further evaluation of abatacept and infliximab while providing no persuasive basis for routine cenicriviroc use.

Should abatacept, cenicriviroc or infliximab be added routinely for hospitalised COVID-19 pneumonia?

ACTIV-1 IM did not establish a routine role for any of the three agents. Abatacept and infliximab produced encouraging secondary mortality signals requiring confirmation; cenicriviroc showed no persuasive benefit.

References

1. RECOVERY Collaborative Group; Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L, et al. Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med* 2021;384(8):693–704.

2. O'Halloran JA, Ko ER, Anstrom KJ, Kedar E, McCarthy MW, Panettieri RA Jr, et al; ACTIV-1 IM Study Group Members. Abatacept, Cenicriviroc, or Infliximab for Treatment of Adults Hospitalized With COVID-19 Pneumonia: A Randomized Clinical Trial. *JAMA* 2023;330(4):328–339.
3. ACTIV-1 IM Study Group. Supplement 2: eTables and eFigures to: Abatacept, Cenicriviroc, or Infliximab for Treatment of Adults Hospitalized With COVID-19 Pneumonia. *JAMA* 2023;330(4):328–339.
4. RECOVERY Collaborative Group. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet* 2021;397(10285):1637–1645.
5. Marconi VC, Ramanan AV, de Bono S, Kartman CE, Krishnan V, Liao R, et al; COV-BARRIER Study Group. Efficacy and safety of baricitinib for the treatment of hospitalised adults with COVID-19 (COV-BARRIER): a randomised, double-blind, parallel-group, placebo-controlled phase 3 trial. *Lancet Respir Med* 2021;9(12):1407–1418.
6. RECOVERY Collaborative Group. Baricitinib in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial and updated meta-analysis. *Lancet* 2022;400(10349):359–368.

MUPIROCIN–IODOPHOR SWAP

Huang SS, Septimus EJ, Kleinman K, Heim LT, Moody JA, Avery TR, et al. Nasal Iodophor Antiseptic vs Nasal Mupirocin Antibiotic in the Setting of Chlorhexidine Bathing to Prevent Infections in Adult ICUs: A Randomized Clinical Trial. JAMA 2023;330(14):1337–1347.

Introduction

Preventing infection in the intensive care unit involves managing patients' own microbial reservoirs as well as limiting transmission between patients. The anterior nares are an important reservoir of *Staphylococcus aureus*. A programme combining nasal decolonisation with antiseptic bathing addresses different anatomical sites, but its effectiveness depends on the products used, their delivery and the susceptibility of the organisms encountered. Replacing one component therefore requires evidence about the whole programme in its intended clinical setting.

The REDUCE MRSA trial established an influential basis for universal decolonisation in adult intensive care units. Its universal strategy combined nasal mupirocin with chlorhexidine bathing and produced larger reductions in bloodstream infection than screening and isolation or targeted decolonisation. This supported treating all patients rather than waiting to identify carriers of methicillin-resistant *S. aureus* (MRSA). It did not establish that alternative nasal agents would preserve the same benefit.¹

Mupirocin is an antibiotic. Repeated population-wide use raises concerns about resistance and the durability of effectiveness. Povidone-iodine is an antiseptic with a different antimicrobial mechanism and offered a plausible alternative. However, changing from an antibiotic to an antiseptic might exchange concerns about resistance for a shorter or less reliable effect on nasal carriage. Experience with brief perioperative nasal preparation could not determine which product would work better during an ICU admission.

A large US hospital system already using universal mupirocin and chlorhexidine provided an opportunity to test a practical policy change. The Mupirocin-Iodophor ICU Decolonization Swap Out Trial asked whether replacing nasal mupirocin with iodophor, while continuing chlorhexidine bathing, would avoid an unacceptable increase in ICU-attributable *S. aureus* clinical cultures. Its non-inferiority design addressed preservation of an established programme's effectiveness, with bloodstream infection providing a separate and clinically important assessment.²

Synopsis

This pragmatic, open-label, hospital-cluster randomised trial involved 137 HCA Healthcare hospitals with 233 adult ICUs in 18 US states. All participating hospitals had adopted universal nasal mupirocin and chlorhexidine bathing in 2013. A 24-month baseline period ran from 1 May 2015 to 30 April 2017, followed by a six-month implementation phase and an 18-month intervention period from 1 November 2017 to 30 April 2019. The implementation phase was excluded from the principal analyses. Funding came from the US Centers for Disease Control and Prevention; Medline donated iodophor. CDC investigators contributed to design, conduct and manuscript review, but not data collection, management or analysis.^{2,3}

Hospitals needed adult ICU services, compatible electronic records and agreement to maintain relevant infection-prevention practices. Important exclusions included a mean ICU stay below two days and systems unable to supply the required data. Of 178 hospitals assessed, 31 were ineligible and ten of the remaining 147 declined. Hospitals were matched using characteristics including volume, infection rates, decolonisation adherence, case mix and estimated baseline mupirocin resistance. Allocation assigned 69 hospitals with 122 ICUs to iodophor and 68 hospitals with 111 ICUs to continued mupirocin. All adult ICUs within a hospital followed the same allocation. Treatment was unblinded, and individual consent was waived for this minimal-risk quality-improvement intervention.

The iodophor policy used 10% povidone-iodine nasal swabs twice daily for five days. The comparator used 2% nasal mupirocin ointment twice daily for five days. Both groups continued daily bathing with 2% chlorhexidine no-rinse cloths throughout the ICU stay, irrespective of MRSA screening status. Existing contact-precaution policies were to remain stable. Implementation included computer-based training, nursing documentation, observed bathing assessments and monthly coaching with adherence feedback. This was a supported hospital-wide change in practice rather than dispensing the alternative product without implementation assistance.

The primary endpoint was the first ICU-attributable *S. aureus* clinical culture, encompassing methicillin-susceptible and methicillin-resistant organisms. Attribution began on the third ICU day and continued through two days after ICU discharge. Surveillance cultures were excluded. Secondary endpoints were MRSA clinical cultures and bloodstream infection from any pathogen within the same attribution window. A bloodstream infection involving a skin commensal required repeated positive cultures within two calendar days. Outcome-specific denominators counted days at risk before a first qualifying event, rather than all occupied ICU days.

The primary analysis used proportional-hazards models with hospital and patient shared frailties to account for clustering and repeated admissions. It compared each group's intervention-period hazard with its own baseline hazard, then compared those changes between groups. The prespecified non-inferiority margin allowed a relative hazard of 1.10, representing at most a 10% deterioration. Power was reported as greater than 80%, with a two-sided significance level of 0.05. Further sample-size assumptions and detailed missing-data procedures were not available in the reviewed report and results supplement; the protocol and statistical analysis plan were excluded from this appraisal as requested.

The main analysis retained hospitals as randomised. Adjusted analyses incorporated demographic characteristics, comorbidity, prior healthcare exposure, surgery, organism history and ICU type. An as-treated analysis required at least two doses of the allocated nasal product and excluded admissions receiving both products. Three hospitals withdrew: two in the mupirocin group lost their adult ICU services, while one iodophor hospital discontinued MRSA contact precautions. Their available data remained in the as-randomised analysis; the as-treated analysis excluded data after withdrawal.

There were 801,668 ICU admissions across the baseline and intervention periods. The intervention period contributed 353,323 admissions: 185,022 in iodophor hospitals and 168,301 in mupirocin hospitals. These were admissions, not an equivalent number of separately randomised individuals. The median age was 66 years, approximately 54% were male, and the median ICU stay was three days. Most units were mixed medical and surgical ICUs, and specialist transplant services were uncommon. The population therefore represented broad community-hospital ICU practice rather than a cohort selected for prolonged critical illness.

Median hospital adherence, defined by receipt of at least two nasal doses, was 79% with iodophor and 89% with mupirocin during the intervention period. Median nasal doses per admission were four and five, respectively. Chlorhexidine adherence was 86% and 88%. These figures describe partial exposure to the programme, not universal completion of five treatment days. Baseline estimates of mupirocin resistance averaged approximately 13% in each group, using available isolates and regional estimates rather than direct testing of every hospital's trial-period organisms.

For *S. aureus* clinical cultures, iodophor hospitals increased from 4.3 to 5.0 events per 1,000 at-risk days, with 3,563 events during the intervention period. Mupirocin hospitals changed from 4.0 to 4.1, with 2,708 intervention-period events. The within-group intervention-versus-baseline hazard ratios were 1.17 (95% confidence interval 1.12–1.23) and 0.99 (0.94–1.04). The relative hazard change was 18.4% greater with iodophor (95%

confidence interval 10.7%–26.6%; $P < 0.001$). Thus, the confidence interval lay beyond the prespecified 10% non-inferiority margin as well as favouring continued mupirocin over the null comparison.

MRSA clinical culture rates changed from 2.1 to 2.3 per 1,000 at-risk days with iodophor and remained approximately 2.0 with mupirocin. Corresponding within-group hazard ratios were 1.13 (1.06–1.20) and 0.99 (0.92–1.06). The relative change was 14.1% greater with iodophor (3.7%–25.5%; $P = 0.007$). All-pathogen bloodstream infection rates remained approximately 2.7 and 2.6 per 1,000 at-risk days, with within-group hazard ratios of 1.00 and 1.01. The between-group comparison was not significant ($P = 0.84$), with an interval extending from about a 9% advantage to an 8% disadvantage.

Adjusted and as-treated analyses continued to favour mupirocin for *S. aureus* clinical cultures; the corresponding relative changes were 20.9% and 27.4% greater with iodophor. Two adverse events were reported in iodophor recipients: mild nasal itching and generalised hives requiring treatment. Both resolved after decolonisation stopped. A post hoc historical comparison with the earlier REDUCE MRSA trial suggested that the mupirocin programme retained substantial effectiveness over time, but this comparison was not randomised within the current study.

Critique

The clearest finding is that switching an established universal decolonisation programme from mupirocin to iodophor worsened its *S. aureus* clinical-culture outcome. This was more than an inconclusive non-inferiority test. The primary confidence interval lay wholly beyond the 10% permitted deterioration, and the comparison also favoured mupirocin against the null. The trial therefore gives a strong practical reason to retain mupirocin in similar programmes when prevention of these cultures is the objective. It does not establish a mortality benefit or an equivalent reduction in clinically adjudicated infection.

Understanding the estimand is essential. The 18.4% result compares changes in hazards between study periods and policies. It is neither an 18.4 percentage-point difference in individual infection risk nor a direct comparison of the two crude intervention-period rates. The model incorporates baseline experience and clustering. The principal Results section describes an 18.4% greater relative hazard change with iodophor, the direction used here. Other parts of the publication describe the same percentage as a reduction with mupirocin, although reciprocal relative changes are not numerically identical. This reporting inconsistency does not reverse the direction or undermine the confidence interval's relationship to the non-inferiority boundary.

Hospital allocation was appropriate for a policy delivered by clinical teams and likely to influence transmission between patients. It limited contamination from simultaneously applying different universal strategies in the same unit. Matching, a two-year baseline, a common electronic-record system and concurrent intervention periods strengthened comparison between hospitals. Retaining data from withdrawing hospitals in the primary analysis also preserved allocation. Nevertheless, hundreds of thousands of admissions should not be mistaken for hundreds of thousands of independent randomised observations. The 137 hospital clusters and the modelling of their variation determine much of the inferential strength.

The baseline comparison improves precision and accommodates pre-existing rate differences, but it introduces dependence on the model used to compare periods. Infection control, case mix and culture collection can change over time. Concurrent randomisation protects against common secular changes; it cannot guarantee that every hospital experiences the same changes. The agreement of adjusted and unadjusted results is reassuring. It does not establish that all relevant differences were measured, particularly where demographic data were missing or additional modelling decisions could not be independently reviewed under the agreed protocol and SAP exclusion.

The primary endpoint is useful for surveillance of programme performance but imperfect as a patient-centred outcome. A clinical culture can reflect infection, colonisation identified during investigation, or differences in decisions to obtain specimens. Excluding surveillance samples prevents routine carriage screening from dominating the result, but it does not convert all positive clinical cultures into confirmed infections. Treatment was visible, and no separate blinded infection-adjudication process was established in the reviewed material. A systematic difference in culture-taking could therefore affect ascertainment. The large, coherent *S. aureus* and MRSA findings remain persuasive, while their clinical meaning should retain the endpoint's actual wording.

Adherence provides a further distinction between product efficacy and policy effectiveness. Receipt of at least two doses was ten percentage points lower with iodophor, and both groups commonly received fewer than the ten planned doses because ICU stays were short. The randomised comparison captures the consequences of deploying each policy under these conditions, including acceptability and implementation. It cannot isolate equal-dose pharmacological effectiveness. The as-treated result does not solve this problem: requiring treatment exposure selects admissions after randomisation and may preferentially exclude short stays, early events or difficult-to-treat patients. Its agreement supports the direction of the policy result without recreating a randomised comparison of fully adherent patients.

The absence of a detectable all-pathogen bloodstream infection difference requires separate interpretation. These events include organisms that nasal *S. aureus* decolonisation may not affect. Both groups received chlorhexidine, which might have maintained protection against other sources and pathogens, but this is a mechanistic explanation rather than a component effect isolated by this trial. The near-null estimate and its interval suggest that a large difference in this endpoint was unlikely under the primary model. They do not show equivalence across all clinical outcomes. Adjusted and as-treated bloodstream-infection intervals also extended beyond a 10% relative disadvantage, discouraging a broad claim that iodophor was proven comparably safe.

Safety reporting was too limited to supply such reassurance independently. Two reported reactions identify possible tolerability problems, including a generalised allergic-type reaction, but routine programme reporting is less sensitive than active systematic ascertainment. The low reported count should not be read as a precise incidence of all adverse effects. Equally, the trial's culture outcome should not be relabelled as evidence of survival benefit: mortality was not the main effectiveness comparison, and descriptive table entries do not provide a valid substitute for a specified clinical outcome analysis.

The historical comparison is the weakest basis for causal claims. Current outcomes were compared with the earlier REDUCE MRSA periods, with only partial overlap between hospitals and substantial elapsed time. Differences in patients, devices, antibiotic use, laboratory practice and other infection-prevention measures could all contribute. The comparison is compatible with continuing benefit from mupirocin-based decolonisation, but cannot demonstrate that resistance failed to emerge. The current trial did not provide serial susceptibility results establishing such a conclusion. Nor can it demonstrate the incremental effect of adding a nasal agent to chlorhexidine: there was no contemporary chlorhexidine-only group.

Finally, applicability is strongest to adult ICUs already operating a mature universal programme with organisational support. A system starting decolonisation from scratch, using a different iodophor formulation or dosing schedule, treating a substantially different resistance ecology, or caring for many prolonged-stay specialist patients might achieve different results. Those possibilities do not negate the observed comparison. They define its boundaries. For comparable hospitals, the alternative antiseptic did not preserve the primary effectiveness of established mupirocin use, even with structured implementation support.

Body of Evidence

The earlier REDUCE MRSA trial explains the foundation of the tested programme. In 43 participating hospitals, three policies compared screening and isolation, screening with targeted decolonisation, and universal decolonisation. During the intervention period, the universal group's MRSA clinical-culture hazard fell by 37% and its all-pathogen bloodstream-infection hazard by 44% relative to its own baseline. These within-group changes should not be quoted as direct treatment hazard ratios against another arm. The reduction in bloodstream infection was greater with the universal strategy than with either alternative. REDUCE MRSA therefore supported a combined universal programme; the current trial addresses which nasal component should be used within it.¹

Evidence for chlorhexidine bathing concerns a related but different intervention. Climo and colleagues conducted a cluster-randomised crossover trial in nine ICU or bone-marrow transplantation units involving 7,727 patients. Hospital-acquired bloodstream infection rates were 4.78 per 1,000 patient-days during chlorhexidine bathing and 6.60 during non-antimicrobial bathing ($P=0.007$). Acquisition of multidrug-resistant organisms also fell. These findings support the plausibility that a shared bathing intervention could influence infection rates in both groups of the Swap Out Trial. They cannot establish that chlorhexidine explains its neutral bloodstream-infection comparison or compensate analytically for the absence of a chlorhexidine-only arm.⁴

Direct mupirocin–iodophor evidence also existed in surgery. Phillips and colleagues randomised patients undergoing arthroplasty or spinal fusion to five days of preoperative nasal mupirocin or brief povidone-iodine application shortly before incision; both groups used chlorhexidine wipes. Deep surgical-site infections occurred after 14 of 855 procedures in the mupirocin group and six of 842 in the iodophor group ($P=0.10$). *S. aureus* deep infections were uncommon. This trial addressed preparation for a defined surgical exposure, with different nasal schedules and a different endpoint. Its non-significant overall result did not prove equal effectiveness, and its findings could not reliably predict sustained ICU decolonisation performance. The later ICU trial supplies the more directly applicable comparison for adult intensive care practice.⁵

Resistance surveillance from REDUCE MRSA provides useful caution about the rationale for switching. Hayden and colleagues tested 3,173 MRSA isolates. Baseline low-level and high-level mupirocin resistance were 7.1% and 7.5%, respectively. The study did not detect a differential increase in resistance between trial arms, but estimates were imprecise and isolates were available from only part of the eligible collection. The findings supported continued surveillance rather than a conclusion that universal

mupirocin could never select resistance. They also concerned an earlier period and cannot replace susceptibility measurements during the Swap Out Trial.⁶

Together, these studies separate three decisions that should not be conflated: whether to implement universal decolonisation, which nasal agent to use, and whether the resulting programme remains effective in a changing microbial environment. REDUCE MRSA informs the first. The Swap Out Trial provides a large, contemporary randomised comparison for the second. Neither eliminates the need to assess the third through local microbiology, adherence and clinically meaningful infection outcomes. Avoiding antibiotic exposure is a reasonable stewardship objective, but it does not by itself establish that an antiseptic substitution preserves benefit.

By the end of 2023, the most directly relevant evidence favoured continued mupirocin with chlorhexidine over the tested iodophor substitution for preventing ICU-attributable *S. aureus* clinical cultures in comparable adult ICUs. The evidence was less decisive for all-pathogen bloodstream infection and did not establish a mortality effect. The appropriate conclusion is therefore specific to the tested programme, product schedules and outcome. Local resistance or intolerance may require a different strategy, but the trial does not identify a subgroup in which routine replacement with this iodophor regimen has equivalent effectiveness.

Should iodophor replace nasal mupirocin in adult ICU universal decolonisation programmes?

The tested iodophor regimen should not routinely replace mupirocin with chlorhexidine bathing in comparable ICUs: it produced worse *S. aureus* clinical-culture outcomes. A difference in all-pathogen bloodstream infection was not established.

References

1. Huang SS, Septimus E, Kleinman K, Moody J, Hickok J, Avery TR, et al. Targeted versus universal decolonization to prevent ICU infection. *N Engl J Med* 2013;368(24):2255–2265.
2. Huang SS, Septimus EJ, Kleinman K, Heim LT, Moody JA, Avery TR, et al. Nasal Iodophor Antiseptic vs Nasal Mupirocin Antibiotic in the Setting of Chlorhexidine Bathing to Prevent Infections in Adult ICUs: A Randomized Clinical Trial. *JAMA* 2023;330(14):1337–1347.

3. Huang SS, et al. Supplementary online content accompanying the Mupirocin-Iodophor ICU Decolonization Swap Out Trial. JAMA. 2023.
4. Climo MW, Yokoe DS, Warren DK, Perl TM, Bolon M, Herwaldt LA, et al. Effect of daily chlorhexidine bathing on hospital-acquired infection. N Engl J Med 2013;368(6):533–542.
5. Phillips M, Rosenberg A, Shopsis B, Cuff G, Skeete F, Foti A, et al. Preventing surgical site infections: a randomized, open-label trial of nasal mupirocin ointment and nasal povidone-iodine solution. Infect Control Hosp Epidemiol 2014;35(7):826–832.
6. Hayden MK, Lolans K, Haffenreffer K, Avery TR, Kleinman K, Li H, et al. Chlorhexidine and Mupirocin Susceptibility of Methicillin-Resistant Staphylococcus aureus Isolates in the REDUCE-MRSA Trial. J Clin Microbiol 2016;54(11):2735–2742.



Trauma Trials

SHINE-TRAUMA

Johansson PI, Fenger Eriksen C, Bovbjerg PE, Gaarder C, Pall M, Henriksen HH, et al. Prostacyclin in trauma patients with hemorrhagic shock: A randomized clinical trial. J Trauma Acute Care Surg 2024;96(3):476–481. Published online 14 November 2023.

Introduction

Control of bleeding and restoration of circulating volume are immediate priorities after major trauma, but survival also depends on limiting subsequent organ injury. Haemorrhagic shock affects the vascular endothelium as well as macroscopic blood flow. Damage to its protective glycocalyx may promote vascular leakage, inflammation and disordered coagulation. A treatment that preserves endothelial function could therefore complement haemorrhage control, provided it does not worsen hypotension or bleeding.

Observational studies supplied a rationale for this approach. In a prospective cohort of 424 trauma patients, catecholamine concentrations were associated with markers of endothelial injury, and both sympathoadrenal activation and endotheliopathy were associated with mortality. These relationships identify a potentially important pathway, but do not establish that modifying an endothelial biomarker will improve recovery. Injury severity can cause both endothelial damage and poor outcome, and a marker of damage is not necessarily a treatment target.¹

Prostacyclin is an endogenous mediator with vasodilator and platelet-inhibitory properties. Experimental work and small clinical studies suggested that low doses might also protect endothelial integrity. Iloprost, a synthetic prostacyclin analogue, offered a way to test this hypothesis. Its familiar vascular effects create a particular tension in haemorrhagic shock: the hoped-for microvascular benefit must be weighed against the possibility of additional haemodynamic instability or impaired haemostasis. Evidence from other forms of critical illness could support investigation, but not routine use in bleeding trauma patients.²

SHINE-TRAUMA evaluated a continuous low-dose iloprost infusion started early after injury and continued for 72 hours. Its principal question was whether patients would spend more of the first 28 days alive and outside intensive care. The report became available online in November 2023, despite its 2024 issue citation, and therefore falls within the CCR24 evidence cut-off. The trial provides a direct clinical test of the endothelial-protection hypothesis in traumatic haemorrhagic shock.

Synopsis

SHINE-TRAUMA was an investigator-initiated, randomised, placebo-controlled, blinded pilot trial conducted at four Scandinavian trauma centres: three in Denmark and one in Norway. Recruitment ran from 24 May 2019 to 14 August 2021. Funding came from Danish Regions Medicinpuljen and the Novo Nordisk Foundation. The available report did not specify additional funder roles in design, analysis or reporting. Iloprost was evaluated as an adjunct to the centres' trauma care, not as a replacement for transfusion or haemorrhage control.²

Eligible patients were at least 18 years old and had clinical signs of haemorrhagic shock, defined by a systolic blood pressure below 90 mmHg or prehospital blood transfusion. The local massive transfusion protocol had to be activated and the first post-admission transfusion started. Randomisation was required within five hours of injury and three hours of emergency-department admission. Important exclusions included withdrawal of active treatment, pregnancy, known iloprost hypersensitivity, severe heart failure classified as New York Heart Association class IV, suspected acute coronary syndrome and estimated weight below 40 kg. Consent for incapacitated patients was obtained through the specified guardian or surrogate arrangements and subsequently confirmed by patients who regained capacity.

A centrally concealed sequence allocated patients 1:1 in variable-sized permuted blocks stratified by centre. Allocation was implemented through the REDCap electronic research database. Iloprost and placebo were similar in colour, consistency and volume. Patients, clinicians, investigators, staff recording outcomes and laboratory results, and statisticians were blinded; treatment identities were withheld until trial procedures and statistical analyses were complete. These arrangements were particularly relevant to outcomes influenced by decisions about continuing intensive care or discharging patients.

The active intervention was intravenous iloprost at 1 ng/kg/min as a continuous infusion for 72 hours. The comparator was an equal-volume saline infusion for the same planned duration. Detailed haemodynamic thresholds for dose interruption or rescue, cumulative delivered doses and a standardised algorithm for other resuscitation treatments were not supplied in the reviewed results materials. Protocols and statistical analysis plans were excluded from this appraisal at the user's request, so such details have not been inferred.

The primary outcome was the number of ICU-free days alive within 28 days of admission, with death during the ICU stay assigned zero. The available description did not fully explain scoring after ICU discharge followed by death or readmission. Secondary

outcomes included all-cause mortality at 28 and 90 days, hospital length of stay, and days free of mechanical ventilation, vasopressors and renal replacement therapy within 28 days. Safety outcomes included serious adverse reactions and selected serious adverse events, particularly ischaemic events. The Methods described ischaemic-event ascertainment within four days, whereas the results table labelled serious-event and reaction outcomes within seven days.

The sample-size calculation sought a 30% increase in ICU-free days, from 15 to 19.5 days. With a two-sided significance threshold of 0.05 and 85% power, 107 patients per group were required. Recruitment stopped after at least 110 patients in each group had started treatment. The primary analysis used linear regression adjusted for centre, with robust standard errors to accommodate non-normal residuals. Binary outcomes used logistic regression; risk ratios and risk differences were additionally derived through g-computation. Iloprost was the reference group, so the displayed adjusted mean differences describe placebo minus iloprost. All reported confidence intervals were 95% intervals.

Of 576 patients screened, 242 were randomised, 123 to iloprost and 119 to placebo. Thirteen were excluded before treatment began. In the iloprost group, two died, one transferred to a ward and one did not have haemorrhagic shock. In the placebo group, five died, three transferred to a ward and one had severe cardiac disease. The population labelled intention-to-treat therefore comprised 119 iloprost and 110 placebo recipients who started intervention, rather than all randomised patients. Three and five patients subsequently discontinued treatment; 116 and 105 contributed to the per-protocol analysis.

Participants were predominantly male and approximately 71% had blunt injury. Median age was 50 years with iloprost and 46 with placebo; median Injury Severity Scores were 22 and 25. Admission median systolic pressures were approximately 100 mmHg in both groups, consistent with eligibility also allowing prior hypotension or prehospital transfusion. Median lactate concentrations were 4.5 and 4.3 mmol/L. Treatment began a median of 3.0 versus 2.7 hours after injury and 1.7 versus 1.6 hours after admission. Admission syndecan-1 concentrations supported the presence of endothelial injury, but a biomarker threshold was not required for entry.

Mean ICU-free days alive within 28 days were 15.64 with iloprost and 13.99 with placebo. The adjusted placebo-minus-iloprost difference was -1.63 days (95% confidence interval -4.64 to 1.38; $P=0.28$). The per-protocol comparison was also non-significant ($P=0.23$). At 28 days, 22 of 117 patients with available data in the iloprost group had died, compared with 21 of 107 in the placebo group: 18.8% versus 19.6%. The adjusted odds ratio, with

iloprost as reference, was 1.01 (0.51–2.00; $P=0.97$). At 90 days, mortality was 23/116 versus 21/104, or 19.8% versus 20.2%, with an odds ratio of 0.98 (0.50–1.93; $P=0.96$).

Ventilator-free days averaged 18.03 versus 16.40, vasopressor-free days 19.86 versus 18.07, and renal-replacement-free days 23.11 versus 22.99. None of these comparisons was statistically significant. Mean hospital stay within the 90-day assessment was 19.96 versus 27.32 days, with an adjusted placebo-minus-iloprost difference of 7.84 days (1.66–14.02; $P=0.01$). Additional analyses examined hospital stay among survivors and after excluding short stays. Exploratory subgroup comparisons considered age, sex, injury severity and mechanism.³

Four of 119 iloprost recipients and five of 108 placebo recipients had a reported serious adverse event in the trial's seven-day window. One iloprost recipient and no placebo recipient had a serious adverse reaction. Transfusion requirements were not significantly different. These observations did not identify an excess of reported events, but the trial had limited precision for uncommon harms and did not collect every possible adverse event.

Critique

The trial did not establish that iloprost increased days alive and free of intensive care. Its point estimate favoured treatment, but the interval included both an important benefit and some disadvantage. With the published placebo-minus-iloprost orientation, the result is compatible with iloprost providing up to approximately 4.6 additional ICU-free days or approximately 1.4 fewer days. This is an interpretation of the reported interval, not a new analysis. A non-significant result does not establish equivalence, and the hospital-stay finding should not displace the primary outcome as the principal test of efficacy.

The most consequential methodological issue is the population called intention-to-treat. Thirteen randomised patients were removed before infusion, including seven who died. The exclusions were more numerous with placebo, and early death is directly relevant to a trial intended to improve survival and organ recovery after shock. Because these deaths preceded drug exposure, they cannot be attributed to pharmacological toxicity or benefit. Nevertheless, excluding them changes the question from the effect of assignment to the effect among patients who survived and remained eligible long enough to start treatment. Blinding reduces the likelihood of deliberate treatment-dependent exclusion, but does not restore the complete randomised comparison.

The extent and consequences of missing outcome data also deserve attention. The mortality denominators were smaller than the treatment-start populations, and follow-up decreased further by 90 days, particularly in the placebo group. Available materials did not give a complete explanation or a corresponding sensitivity analysis for these missing observations. With only 43 observed deaths at 28 days and virtually identical reported mortality proportions, a small number of unobserved outcomes could matter. The appropriate response is to preserve the actual denominators and acknowledge uncertainty, rather than describe mortality follow-up as complete or impute survival without evidence.

The trial's concealment and blinding are concrete strengths. Central allocation, centre stratification and an indistinguishable infusion reduced opportunities to select treatment or alter care according to assignment. Blinded outcome collection and analysis further limited ascertainment and analytical bias. These features strengthen confidence that the observed differences were not simply products of knowing which infusion patients received. They remain important even though the primary comparison was inconclusive and some post-randomisation exclusions weakened the analysis.

The chosen endpoint combines survival and resource use, but its precise construction matters. Assigning zero ICU-free days after death in intensive care avoids treating a short terminal admission as a favourable result. However, the reported definition leaves uncertainty about deaths after discharge and recurrent ICU admissions. Those situations can change the interpretation of a days-free endpoint. Moreover, time in intensive care depends on local discharge arrangements and the nature of the injury, as well as organ function. Blinding and adjustment for centre help, but do not make this a direct measurement of endothelial repair.

The anticipated treatment effect was ambitious. A gain of 4.5 ICU-free days represents a substantial change in recovery, and the trial was designed to detect that scale of benefit. It was not sized to reliably establish a smaller but clinically worthwhile effect or a mortality difference. Robust standard errors were a reasonable response to the bounded, non-normal outcome distribution, but the report did not provide all the distributional or planning assumptions required for independent reconstruction. The label "pilot" is therefore consistent with the residual uncertainty, rather than a reason to reinterpret a non-significant result as evidence of benefit.

Patient selection was clinically practical but did not specifically identify patients most likely to respond to endothelial treatment. Entry depended on shock and transfusion criteria, not a syndecan-1 threshold. Elevated median syndecan-1 supports the biological rationale at a group level, yet does not demonstrate that every participant had the same

mechanism or a reversible target. Blunt and penetrating injuries, associated brain injury and different anatomical damage can produce very different recovery trajectories. No demonstrated interaction identifies a responsive subgroup, and the trial does not establish that earlier, longer or higher-dose administration would improve outcomes.

The shorter hospital stay is potentially meaningful, but requires confirmation. It was one of several secondary comparisons without a demonstrated multiplicity-protected hierarchy in the reviewed material. Length of stay is influenced by death, discharge destination, rehabilitation capacity and the injuries requiring continued treatment. The supplementary analyses among survivors and patients staying beyond three or seven days explore alternative explanations, but condition on events after randomisation. They cannot prove that survival or early discharge played no causal role. Their P values also arise from different subsets and analyses and should not be treated as independent replication of the main adjusted mean difference.

The exploratory subgroup discussion is especially vulnerable to overinterpretation. A comparison approaching conventional significance in patients with an Injury Severity Score of at least 25 does not establish that severity modifies the treatment effect. The supplied table presents numerous within-subgroup comparisons across several outcomes, without a clear interaction analysis. Its mean ICU-free days in the higher-severity subgroup were 10.93 and 7.33, whereas the main text described 10.5 and zero without clearly reconciling the summary measures. Those figures should not be combined into an apparently precise treatment effect or used to select patients for clinical treatment.

Finally, the safety conclusion should be narrower than the publication's assertion that iloprost was safe. Similar transfusion requirements and a small number of reported ischaemic events are reassuring observations, particularly for a vasodilating, platelet-inhibiting drug. They do not exclude clinically important bleeding, hypotension or uncommon harm. Safety ascertainment was restricted, its stated window differed between Methods and Table 2, and there was one serious adverse reaction with iloprost. The absence of a statistically significant difference in such a small dataset supports further monitored research, rather than routine adjunctive administration after traumatic haemorrhage.

Body of Evidence

The observational foundation is relevant but cannot supply the treatment effect missing from the trial. Johansson and colleagues' 424-patient cohort linked adrenaline, syndecan-1 and other endothelial markers with injury severity and mortality. Adjustment

for measured clinical factors strengthened the prognostic association. It did not remove every possible common cause or demonstrate that prostacyclin would reverse the process. SHINE-TRAUMA was therefore an important transition from describing a high-risk biological phenotype to testing a therapeutic intervention. Its inconclusive result shows why prognostic evidence and evidence of therapeutic responsiveness must remain separate.¹

The CO-ILEPSS trial investigated a related approach in septic shock. Twenty-four patients received combined iloprost and eptifibatide or placebo, with 18 completing the seven-day period used for the primary per-protocol biomarker analyses. Direct between-group comparisons did not show differences in the primary outcomes. Some biomarkers and organ-failure scores changed significantly within the active group, but a change in one group and no significant change in another is not itself proof of a treatment difference. The trial also tested two drugs together in a different disease. It offered exploratory biological and tolerability information, not a reliable estimate of isolated iloprost efficacy in traumatic haemorrhage.⁴

A subsequent trial in 80 mechanically ventilated patients with COVID-19 used the same 1 ng/kg/min prostacyclin dose for 72 hours, but required an elevated thrombomodulin concentration for entry. Median days alive without ventilation were approximately 16 versus five, with a reported difference of 10.96 days and a 95% confidence interval from -5 to 21 days ($P=0.07$). The primary outcome was therefore inconclusive. A lower mean organ-failure score was a secondary finding. This trial supplied a reason for further investigation, while its biomarker selection, pulmonary disease and later phase of critical illness made direct extrapolation to acute traumatic shock uncertain.⁵

Chronology also matters. SHINE-TRAUMA began recruitment in May 2019, before either the published CO-ILEPSS results or the COVID-19 trial report was available. These studies contribute to interpretation at the end of 2023, but should not be described as established positive evidence that justified the original design. The biological case for treatment preceded them; their clinical results remained exploratory. Repeated favourable point estimates or secondary outcomes across small studies do not automatically constitute a confirmed patient-centred benefit, particularly when populations, co-interventions and primary endpoints differ.

The available programme of research leaves several connected questions. An endothelial biomarker may identify risk without identifying a modifiable mechanism. An infusion may change endothelial biology without improving survival or recovery. A reduction in hospital stay may be valuable without proving either of those mechanisms. A future confirmatory trial would need to distinguish these possibilities through a clearly

specified clinical endpoint, retention of all randomised patients, transparent missing-data handling and systematic safety collection. Any proposal to enrich recruitment by injury severity or an endothelial marker would require prospective justification rather than reliance on an isolated exploratory P value.

By the end of 2023, SHINE-TRAUMA supported the feasibility of studying early low-dose iloprost in selected patients receiving modern trauma care. It did not establish improved ICU-free survival, a mortality benefit or comprehensive safety. The shorter hospital stay is a credible hypothesis for confirmation, but not sufficient evidence to make iloprost routine treatment for haemorrhagic shock. The central clinical implication is continued uncertainty about benefit, with enough biological rationale and preliminary experience to support appropriately designed research.

Should low-dose iloprost be added routinely after traumatic haemorrhagic shock?

No. SHINE-TRAUMA did not establish improved ICU-free days or survival; the shorter hospital stay remains a secondary finding requiring confirmation, and safety is not definitively established.

References

1. Johansson PI, Henriksen HH, Stensballe J, Gybel-Brask M, Cardenas JC, Baer LA, et al. Traumatic Endotheliopathy: A Prospective Observational Study of 424 Severely Injured Patients. *Ann Surg* 2017;265(3):597–603.
2. Johansson PI, Fenger Eriksen C, Bovbjerg PE, Gaarder C, Pall M, Henriksen HH, et al. Prostacyclin in trauma patients with hemorrhagic shock: A randomized clinical trial. *J Trauma Acute Care Surg* 2024;96(3):476–481. Published online 14 November 2023.
3. Johansson PI, et al. Supplementary Table S1 and Supplementary Figure 1 accompanying SHINE-TRAUMA. *J Trauma Acute Care Surg*. 2023 (online supplementary material).
4. Berthelsen RE, Ostrowski SR, Bestle MH, Johansson PI. Co-administration of iloprost and eptifibatide in septic shock (CO-ILEPSS)—a randomised, controlled, double-blind investigator-initiated trial investigating safety and efficacy. *Crit Care* 2019;23:301.

5. Johansson PI, S e-Jensen P, Bestle MH, Clausen NE, Kristiansen KT, Lange T, et al. Prostacyclin in Intubated Patients with COVID-19 and Severe Endotheliopathy: A Multicenter, Randomized Clinical Trial. *Am J Respir Crit Care Med* 2022;205(3):324–329.

CRYOSTAT-2

Davenport R, Curry N, Fox EE, Thomas H, Lucas J, Evans A, et al; for the CRYOSTAT-2 Principal Investigators. Early and Empirical High-Dose Cryoprecipitate for Hemorrhage After Traumatic Injury: The CRYOSTAT-2 Randomized Clinical Trial. JAMA 2023;330(19):1882–1891.

Introduction

Major traumatic haemorrhage requires rapid control of bleeding and restoration of effective coagulation. Fibrinogen is the substrate for fibrin formation, and its concentration can fall through consumption, fibrinolysis and dilution. Plasma supplies multiple clotting factors, but concentrated fibrinogen replacement may be needed when fibrinogen becomes deficient. The clinical uncertainty is not simply whether fibrinogen matters: it is which patients require additional replacement, how much they should receive and how quickly it must be delivered.¹

Cryoprecipitate is a concentrated blood component used for fibrinogen replacement in UK and US trauma practice. It is commonly incorporated into later packs of a major haemorrhage protocol. Earlier empirical administration could correct a developing deficiency before laboratory results become available. It could also expose patients whose fibrinogen is adequate to additional blood products without benefit. Preparation and transport of a frozen component introduce practical delays during a period in which bleeding and physiological deterioration can evolve rapidly.

The original CRYOSTAT feasibility trial showed that earlier cryoprecipitate administration was achievable in two UK trauma centres and increased fibrinogen concentrations during active bleeding. Its small sample could not establish a survival advantage. A larger trial was needed to determine whether moving concentrated fibrinogen replacement forward in the resuscitation sequence improved patient outcomes when added to contemporary, balanced blood-component therapy.²

CRYOSTAT-2 therefore tested a policy of giving an additional high dose of cryoprecipitate empirically to severely injured patients requiring major haemorrhage protocol activation. It did not require laboratory confirmation of low fibrinogen, and standard care could itself include cryoprecipitate. The central question was whether this extra early dose improved survival across the eligible bleeding population. Distinguishing that policy from targeted treatment of confirmed deficiency is essential to interpreting the trial.

Synopsis

CRYOSTAT-2 was a phase 3, open-label, individually randomised, parallel-group trial conducted at 26 major trauma centres between August 2017 and November 2021. Twenty-five centres were in the UK and one was in the US. The UK National Institute for Health and Care Research Health Technology Assessment programme and Barts Charity funded the study. The funder had no role in conduct, analysis, interpretation, manuscript preparation or the decision to publish. Independent monitoring and steering committees provided oversight.^{1,3}

Patients were eligible if they were at least 16 years old, or judged to be adults when identity was unknown, with severe injury and active bleeding requiring activation of the local major haemorrhage protocol. At least one unit of a blood component had to have been started or given. Protocol activation included systolic pressure below 90 mmHg at some point, rather than necessarily at randomisation. Interhospital transfers, injuries judged incompatible with life and presentation more than three hours after injury were excluded. Emergency consent waivers permitted enrolment, with consent or representative agreement sought subsequently for continued data collection.

Allocation was 1:1 through sequential opaque sealed envelopes held at each centre. The computer-generated sequence used undisclosed varying block sizes and stratification by centre. Treatment was unblinded. Both groups received the local major haemorrhage protocol, reviewed for consistency with relevant guidance. Typical packs contained four red-cell units and four fresh-frozen plasma units, with platelet pools introduced in the second and subsequent packs to approach balanced component replacement. Two pools of cryoprecipitate, approximately 4 g fibrinogen equivalent, were also typically included in the second and subsequent packs.

The intervention added three early cryoprecipitate pools, equivalent to approximately 6 g fibrinogen, to that standard care. A UK pool represented five individual cryoprecipitate donations. The detailed Methods specified starting the additional treatment as soon as possible and within 90 minutes of hospital admission. The abstract instead used 90 minutes from randomisation; this chapter follows the Methods and the admission-based delivery analyses. The strategy did not require a fibrinogen measurement immediately before treatment. Other blood products and haemorrhage management continued according to clinical need.

The primary endpoint was all-cause mortality at 28 days. Secondary outcomes included mortality at six hours, 24 hours, six months and 12 months; deaths attributed to bleeding; 24-hour transfusion requirements; critical care and hospital stay; discharge destination; and functional and quality-of-life measures. Transfusion totals included

prehospital products and the study cryoprecipitate. Symptomatic venous and arterial thrombotic events were collected to day 28 or discharge, whichever occurred first. Stay and ventilation outcomes used discharge or extubation as the event and death as a competing risk. Massive transfusion was added as a post hoc outcome.

The trial sought a seven percentage-point absolute mortality reduction from an assumed 26%, with 90% power and an overall two-sided significance level of 5%. A group-sequential design used O'Brien–Fleming stopping guidance after 500 and 1,000 patients had completed 28-day follow-up. The final significance threshold was $P < 0.0453$. The initial target of 1,568 participants allowed 2.5% attrition; this was increased to 1,600 when the expected dropout allowance rose to 4.4%. These design details are stated in the main report. The SAP was subsequently checked specifically to clarify the intervention's 90-minute treatment window.

The primary model was mixed logistic regression adjusted for centre. Additional analyses included adjustment for clinical risk factors, an unadjusted model, and a sensitivity analysis changing 2% of outcomes inferred from discharge to death. Patients discharged before day 28 were otherwise inferred to be alive at that point. Missing covariates could be multiply imputed, but missing primary and secondary outcomes were not imputed. There was no adjustment for multiple comparisons among secondary and subgroup analyses. A planned per-protocol analysis excluded selected ineligible patients and those dying within 90 minutes, but did not simply remove everyone failing to receive the intervention.

Of 9,036 patients assessed, 2,756 were eligible and 1,604 were randomised: 799 to additional cryoprecipitate and 805 to standard care. Research-team unavailability accounted for 888 of the 1,152 eligible patients not enrolled. Primary outcome data were available for 760 and 771 patients, respectively; 73 were missing, most commonly because of withdrawal of consent. The median time from admission to randomisation was 15 minutes. Median age was 39 years, 79% were male, median Injury Severity Score was 29, and 36% had penetrating injuries. Approximately one-quarter had severe head injury. Only one-third remained hypotensive on hospital arrival. Prehospital blood transfusion had been given to 43%, and 96% received tranexamic acid before or after arrival.

Cryoprecipitate was given within 24 hours to 665/785 patients with exposure data in the intervention group and 256/795 in standard care, or 85% versus 32%. Median time to first administration among recipients was 68 versus 120 minutes. The flow diagram identified 434 intervention patients as meeting the full assigned intervention requirements; delays, reduced doses, cessation of bleeding and death contributed to

non-delivery. Median cryoprecipitate exposure was three pools versus zero, while median red-cell and plasma exposure was five and four units in both groups. Median total blood-product use was 12 versus ten units, including the added cryoprecipitate.

At day 28, 192/760 intervention patients and 201/771 standard-care patients had died: 25.3% versus 26.1%. The centre-adjusted odds ratio was 0.96 (95% confidence interval 0.75–1.23; $P=0.74$). Six-hour mortality was 7.1% versus 8.6% and 24-hour mortality 11.2% versus 12.2%, without significant differences. Longer-term mortality, critical care and hospital stays, and measured functional outcomes also showed no demonstrated benefit. The estimated cumulative incidence of thrombotic events by day 28 was 12.7% versus 12.9%, a difference of -0.2 percentage points (-4.1 to 3.7 ; $P=0.89$).

The per-protocol primary comparison was 163/706 versus 154/683 deaths, with an odds ratio of 1.03 (0.77–1.37; $P=0.83$). The clinically risk-adjusted sensitivity analysis gave an odds ratio of 1.15 (0.93–1.42; $P=0.20$). In the prespecified penetrating-injury subgroup, mortality was 45/277 versus 27/271, or 16.2% versus 10.0%, with an odds ratio of 1.74 (1.20–2.51). For blunt injury it was 30.4% versus 34.8%, with an odds ratio of 0.82 (0.62–1.09). The treatment-by-injury-type interaction had $P=0.004$. Analyses by achieved cryoprecipitate timing showed varying associations, but timing itself was not randomised.

Critique

CRYOSTAT-2 does not support routinely adding three early cryoprecipitate pools to a major haemorrhage protocol for every eligible bleeding trauma patient. The primary estimate was close to no effect, and no coherent improvement emerged in early mortality, subsequent recovery or other component requirements. This conclusion concerns the empirical policy as delivered. It does not establish that correcting measured hypofibrinogenaemia is ineffective, nor that cryoprecipitate and fibrinogen concentrate are interchangeable. Neither of those comparisons was randomised.

The achieved treatment contrast was substantial but incomplete. Most intervention patients received cryoprecipitate, whereas about one-third of controls did so, and first administration occurred earlier in the intervention group. However, only 434 of 799 allocated patients met the full intervention requirements recorded in the flow diagram. Some no longer needed additional haemostatic treatment, some received an incomplete dose and others received it too late. This limits a claim about the biological efficacy of perfectly timed replacement. It also reveals the practical effectiveness of trying to introduce the policy across a trauma system, which is directly relevant to implementation.

The standard-care exposure is not simply contamination to be removed analytically. Cryoprecipitate was already an accepted component of the comparator protocol, and the trial asked about adding an early empirical dose. Removing controls who subsequently received it would select patients according to bleeding severity and subsequent treatment. The absence of a reduction in red-cell or plasma use, despite the additional concentrated product, is clinically informative. Total product exposure increased; it would be misleading to summarise the trial as showing no difference in every transfusion measure.

Timing is biologically plausible but cannot rescue the negative policy result through subgroup analysis. Patients receiving cryoprecipitate within 45 minutes were more severely injured and shocked than those treated later. Their higher mortality may reflect those differences, while lower mortality in the 61–90-minute category may reflect survival and clinical selection up to that time. Each category was compared with the whole standard-care group, not with a corresponding randomised timing group. These analyses cannot identify an optimal minute for treatment or establish harm from prompt delivery. They instead illustrate why treatment timing requires a design that preserves a valid comparison.

The population was sufficiently ill to test an important clinical question: control mortality closely matched the planning assumption, injury severity was high and many patients needed substantial transfusion. Nevertheless, severe injury and prior hypotension are not equivalent to ongoing fibrinogen deficiency. By admission, most patients were no longer below the systolic pressure threshold, and some ceased bleeding before treatment could be given. High use of tranexamic acid and prehospital blood products may also have altered the coagulation state. Without a required pre-treatment fibrinogen measurement, the trial cannot identify which patients had a deficit, whether it was corrected, or whether adequate concentrations were unnecessarily increased.

All-cause mortality was a strong primary endpoint because it captures both benefit and harm without depending on attribution. It also includes deaths that additional fibrinogen may not plausibly prevent. In the supplement, uncontrolled bleeding accounted for 46 and 41 deaths, while traumatic brain injury accounted for 67 and 78. This can reduce sensitivity to a haemostatic effect but is not a reason to ignore the patient-centred outcome. Early mortality and the severe-head-injury subgroup did not establish a benefit either. Restricting attention after the fact to a seemingly responsive cause of death would weaken the randomised inference.

Missing primary outcomes remain a limitation. Approximately 5% of participants were absent from the 28-day binary analysis, predominantly after consent withdrawal.

Randomisation cannot guarantee an unbiased complete-case comparison if missingness relates jointly to treatment and prognosis. Inferring that patients discharged before day 28 remained alive is a separate assumption. The sensitivity analysis changing a small proportion of these inferred survivors to deaths was reassuring for that assumption, but did not address every possible pattern among the 73 patients with unavailable outcomes. The chapter therefore retains the analysed denominators rather than describing complete follow-up.

Adjustment also adds useful context. The primary centre-adjusted result mildly favoured cryoprecipitate, whereas additional clinical risk adjustment shifted the point estimate towards harm. Both intervals crossed the null. This does not justify replacing the prespecified primary analysis with whichever result is preferred; it shows that the small apparent overall advantage was not robust evidence of benefit. Baseline neurological and injury-severity differences, incomplete covariate data and the chosen adjustment model may affect estimates. Across these analyses, the consistent conclusion is non-establishment of improved survival, with residual uncertainty about smaller effects.

The penetrating-injury result deserves attention because it was prespecified and supported by an interaction test. It is stronger evidence of possible heterogeneity than merely finding a significant result in one subgroup and not another. Nevertheless, multiple comparisons were unadjusted, subgroup event counts were limited, and a specific mechanism of harm was not demonstrated. The overall thrombotic-event comparison was neutral, but that does not rule out subgroup harm or establish that the mortality difference was caused by thrombosis. The finding argues against indiscriminate empirical treatment, while remaining insufficient to define a proven beneficial strategy for blunt trauma.

External validity is strongest for well-developed UK trauma systems. Only 49 patients were recruited in the US, where prehospital practice, transport intervals and available blood products differed. Many otherwise eligible patients were missed because research staff were unavailable, and patients transferred between hospitals were excluded. These features constrain extrapolation to different trauma pathways. They do not negate the main result: across the participating centres, a broadly applied additional early cryoprecipitate policy increased exposure without demonstrating improved survival. Further work should test a clearly defined alternative, rather than assume that changing timing, dose or patient selection will necessarily reverse that finding.

Body of Evidence

The original CRYOSTAT study randomised 43 patients to standard haemorrhage treatment or two additional early cryoprecipitate pools. In the early-treatment group, 85% received cryoprecipitate within 90 minutes; median first administration was 60 versus 108 minutes. Fibrinogen concentrations were higher and remained above 1.8 g/L during active haemorrhage. Mortality was not significantly different. These findings established feasibility and a biological effect, not survival efficacy. CRYOSTAT-2 increased the dose, expanded delivery across many centres and addressed the mortality question directly. Its outcome therefore carries substantially greater weight for the routine empirical policy.²

E-FIT 1 investigated whether fibrinogen concentrate could solve the delivery problem. This blinded UK pilot compared 6 g concentrate with placebo alongside standard major haemorrhage care. Only 27 of 39 participants with evaluable timing data received the study intervention within 45 minutes, below the trial's feasibility criterion. Concentrations increased more with fibrinogen, but clinical outcomes did not establish benefit. The relevant lesson is that a concentrated product alone does not guarantee rapid delivery: recognition, randomisation, preparation and bedside administration all contribute. It also reinforces the distinction between increasing a laboratory concentration and improving an outcome that matters to patients.⁴

FEISTY compared fibrinogen concentrate with cryoprecipitate in an Australian pilot using viscoelastic evidence of hypofibrinogenaemia to guide replacement. One hundred patients were randomised and 62 received replacement after a qualifying deficit was identified. Median time to delivery was 29 minutes with concentrate and 60 minutes with cryoprecipitate, and fibrinogen increased with both products. Its primary endpoint was treatment timing, and it was not a definitive survival comparison. FEISTY therefore informs product logistics in selected deficient patients; it cannot establish that concentrate would improve mortality if substituted into the empirical CRYOSTAT-2 policy.⁵

ITACTIC addressed another part of the pathway: how to guide haemostatic treatment added to an empirical major haemorrhage protocol. Among 396 patients, viscoelastic-assay guidance did not improve the proportion alive and free of massive transfusion at 24 hours compared with conventional coagulation-test guidance, 67% versus 64% (odds ratio 1.15; 95% confidence interval 0.76–1.73). Both groups received active haemostatic management. This does not make diagnostic information irrelevant, but it prevents the CRYOSTAT-2 result being used to claim that an alternative test-guided strategy is already proven to improve survival. A direct test of targeted fibrinogen replacement remains a different question.⁶

These studies address related but distinct decisions. CRYOSTAT examined feasibility, E-FIT examined early concentrate delivery, FEISTY compared replacement products in patients with a demonstrated deficit, and ITACTIC compared guidance strategies within active resuscitation. Their differences in eligibility, product exposure and primary outcome prevent a simple narrative in which all early fibrinogen studies support the same treatment. The earlier feasibility study was available when CRYOSTAT-2 was designed; E-FIT, FEISTY and ITACTIC contribute additional context available by the end of 2023.

The accumulated evidence supports a restrained interpretation. Fibrinogen is central to clot formation, and concentrated products can raise its level, but empirical supplementation across a broad bleeding population did not improve survival in the largest directly relevant trial. There remains uncertainty about rapid replacement in patients with an objectively demonstrated deficit, different delivery systems and the reason for the penetrating-injury signal. Those questions should guide future comparisons. They do not provide evidence to add the tested extra dose routinely to otherwise appropriate major haemorrhage care.

Should all bleeding trauma patients receive an additional early high dose of cryoprecipitate?

CRYOSTAT-2 does not support routine empirical addition of three cryoprecipitate pools to standard major haemorrhage care. Targeted correction of confirmed fibrinogen deficiency was not the comparison tested.

References

1. Davenport R, Curry N, Fox EE, Thomas H, Lucas J, Evans A, et al; for the CRYOSTAT-2 Principal Investigators. Early and Empirical High-Dose Cryoprecipitate for Hemorrhage After Traumatic Injury: The CRYOSTAT-2 Randomized Clinical Trial. *JAMA* 2023;330(19):1882–1891.
2. Curry N, Rourke C, Davenport R, Beer S, Pankhurst L, Deary A, et al. Early cryoprecipitate for major haemorrhage in trauma: a randomised controlled feasibility trial. *Br J Anaesth* 2015;115(1):76–83.
3. Davenport R, et al. Supplement 3: supplemental online content accompanying CRYOSTAT-2. *JAMA*. 2023.

4. Curry N, Foley C, Wong H, Mora A, Curnow E, Zarankaite A, et al. Early fibrinogen concentrate therapy for major haemorrhage in trauma (E-FIT 1): results from a UK multi-centre, randomised, double blind, placebo-controlled pilot trial. *Crit Care* 2018;22:164.
5. Winearls J, Wulschleger M, Wake E, McQuilten Z, Reade M, Hurn C, et al. Fibrinogen Early In Severe Trauma study (FEISTY): results from an Australian multicentre randomised controlled pilot trial. *Crit Care Resusc* 2021;23(1):32–46.
6. Baksaas-Aasen K, Gall LS, Stensballe J, Juffermans NP, Curry N, Maegele M, et al. Viscoelastic haemostatic assay augmented protocols for major trauma haemorrhage (ITACTIC): a randomized, controlled trial. *Intensive Care Med* 2021;47(1):49–59.
7. CRYOSTAT-2 investigators. Statistical analysis plan, version 2, 20 May 2022. [ISRCTN14998314](#).

Possible Reporting Inconsistencies. The abstract places the 90-minute cryoprecipitate target after randomisation, whereas the Methods and SAP place it after hospital admission. These are different treatment windows. This chapter uses the hospital-admission clock specified in the Methods and SAP.^{1,7}

UK-REBOA

Jansen JO, Hudson J, Cochran C, MacLennan G, Lendrum R, Sadek S, et al; UK-REBOA Study Group. Emergency Department Resuscitative Endovascular Balloon Occlusion of the Aorta in Trauma Patients With Exsanguinating Hemorrhage: The UK-REBOA Randomized Clinical Trial. JAMA. 2023;330(19):1862–1871.

Introduction

A patient with uncontrolled torso haemorrhage can deteriorate faster than definitive surgery or embolisation can be arranged. Resuscitative endovascular balloon occlusion of the aorta (REBOA) offers an intuitively attractive bridge: a balloon introduced through the femoral artery temporarily reduces blood flow to the bleeding region while supporting the circulation above it. The central question is whether this manoeuvre buys useful time or adds a hazardous procedure to an already time-critical resuscitation.¹

The physiological bargain is substantial. Aortic occlusion sacrifices distal perfusion, and subsequent reperfusion can produce further instability. Femoral access may be difficult in profound shock. Preparation, cannulation and balloon positioning also occupy staff and time that might otherwise advance definitive haemostasis. The clinically relevant intervention therefore includes the decision to attempt REBOA and its consequences for the surrounding trauma pathway, as well as the effect of an inflated balloon.¹

By 2019, a joint US professional statement recognised the absence of high-grade evidence for improved survival and emphasised integration with rapid definitive haemorrhage control. It treated REBOA as a temporary component of a wider trauma system, with potentially serious vascular and ischaemic complications. This caution reflected a difficult evidence base: clinicians preferentially select the procedure for patients whose bleeding, physiology and response to resuscitation already distinguish them from possible controls.²

UK-REBOA addressed that uncertainty through emergency randomisation in established major trauma centres. Its hypothesis was that adding a REBOA strategy to standard care would improve survival after life-threatening torso haemorrhage. The trial was designed to evaluate effectiveness within a functioning trauma service, where some patients recover before inflation and others deteriorate before access is achieved. That distinction between a treatment strategy and successful device deployment is essential to interpreting its findings.¹

Synopsis

UK-REBOA was a pragmatic, open-label, Bayesian, group-sequential, registry-enabled randomised trial. Sixteen UK major trauma centres participated, although twelve enrolled patients. Recruitment ran from 30 October 2017 to 16 March 2022. The National Institute for Health and Care Research Health Technology Assessment programme funded the trial. The report describes consultation with the funder and external reviewers when revising aspects of the statistical design.^{1,3}

Eligible patients were aged, or believed to be aged, at least 16 years and had confirmed or suspected life-threatening torso haemorrhage judged amenable to adjunctive REBOA. Known or suspected pregnancy and clearly unsurvivable injuries were exclusions. Eligibility depended on clinical judgement rather than a mandatory numerical blood-pressure threshold. Patients unable to consent were enrolled under emergency research provisions, with consent for continued participation sought once the immediate life-threatening condition had resolved.^{1,3}

A web-based system, accessible by mobile phone, allocated patients equally using random blocks of two or four. Clinicians could not be blinded. Both groups received standard trauma care, including airway management, balanced blood-product resuscitation, tourniquets where appropriate and operative or endovascular haemorrhage control. Open aortic occlusion remained permissible. The intervention added a strategy of REBOA, with device choice, balloon position and use of partial occlusion determined clinically. Inflation could occur in the descending thoracic aorta, zone I, or the distal abdominal aorta above its bifurcation, zone III.¹

Sites received a structured training programme, initially delivered over two days and subsequently compressed into one extended day after four courses. It included equipment familiarisation, ultrasound-guided femoral access using perfused simulators, team scenarios and implementation support. Refresher training was encouraged. The appendix describes the curriculum and support arrangements but does not quantify every operator's previous clinical experience or maintained procedural proficiency.³

The primary outcome was all-cause mortality at 90 days. Registry data, trial records and national outcome information supported ascertainment. Secondary outcomes included mortality at other time points, definitive haemorrhage-control procedures and their timing, blood-product use, complications, and hospital- and intensive-care-unit-free days to day 90. Patients dying by day 90 received zero free days. Three-hour mortality was added before recruitment ended; it was not part of the original outcome set. Definitive haemorrhage control comprised specified operations or embolisation, with timing measured to procedure commencement.^{1,3}

The target of 120 participants reflected estimated recruitment feasibility. Interim analyses were planned after 40 and 80 participants. The primary intention-to-treat Bayesian logistic model used a minimally informative normal prior for the treatment log odds ratio, centred at zero with standard deviation 1.28. An odds ratio above one indicated increased mortality with REBOA. The report acknowledges an initial variance-calculation error that overstated the design's operating characteristics; revisions relaxed the success threshold and added informative-prior analyses. The reported primary result used the minimally informative prior.¹

Recruitment stopped following the 80-patient analysis, when the posterior probability of increased mortality was 90.1%, exceeding the 90% harm boundary. Ten additional patients had already been enrolled while outcome data accrued, giving 46 assigned to REBOA and 44 to standard care. One control participant withdrew consent for further data collection on day four, leaving 89 patients in the 90-day analysis. The median Injury Severity Score was 41 in both groups, and 97% had blunt injuries. Median ages were 46 and 39 years. Admission systolic pressure was lower with REBOA, 84 versus 99 mm Hg; pressures of 70 mm Hg or less occurred in 18/44 versus 9/42 patients with measurements.¹

Only 19/46 patients assigned to REBOA had a balloon inflated. Seventeen improved sufficiently that inflation was no longer required, eight had unsuccessful arterial access, and two deteriorated before access. Two control patients received REBOA. Among the 19 allocated patients undergoing inflation, ten received zone I and nine zone III occlusion; eight had partial occlusion. Median time from emergency-department arrival to inflation was 32 minutes, and median inflation duration was 29 minutes, with an interquartile range of 19–64 minutes. Arrival-to-inflation time was not a direct measurement of cannulation duration.¹

At 90 days, 25/46 REBOA patients died, compared with 18/43 controls, 54% versus 42%. The odds ratio was 1.58, with a 95% credible interval of 0.72–3.52, and an 86.9% posterior probability of increased mortality. The modelled absolute difference was 11.3 percentage points, with a credible interval from –8.1 to 30.1 percentage points. The posterior probability of a benefit as large as an odds ratio of 0.77 or lower was 3.7%. Covariate adjustment gave an odds ratio of 1.80, credible interval 0.59–5.59, with an 84.9% probability of increased mortality.^{1,3}

Earlier mortality also favoured standard care. Death within 24 hours occurred in 17/46 versus 10/44 patients, odds ratio 1.85, credible interval 0.79–4.46. Six-hour mortality was 13/46 versus 4/44, odds ratio 3.14, credible interval 1.13–9.76; three-hour mortality was 11/46 versus 2/44, odds ratio 4.25, credible interval 1.33–15.99. The corresponding

probabilities of increased mortality were 91.8%, 98.6% and 99.3%. In-hospital and six-month mortality reproduced the 90-day death counts. Bleeding accounted for eight REBOA deaths and three control deaths, most within the first day.^{1,3}

Definitive haemorrhage-control procedures were recorded in 14/46 versus 19/44 patients. Median randomisation-to-procedure time was 83 versus 64 minutes, with timing available for twelve REBOA recipients of such procedures. Mean hospital-free days were 22 versus 41, a modelled difference of –18.58 days, credible interval –32.86 to –3.93. Intensive-care-free days and blood-product requirements showed no clear improvement. Median red-cell use over 24 hours was seven versus nine units, with an incidence rate ratio of 0.92, credible interval 0.66–1.29.^{1,3}

Complications were reported in six REBOA patients and ten controls; no serious adverse device events were reported. Recorded events nevertheless included vascular-access complications, distal embolism, limb ischaemia and renal replacement therapy. The small numbers did not establish equivalence of safety. Supplementary analyses examined actual treatment delivery and the influence of excluding early site cases, but these were exploratory departures from the full randomised comparison.^{1,3}

Critique

The principal finding is a concerning absence of benefit from introducing the REBOA strategy, with mortality estimates consistently favouring standard care. It is neither a precise demonstration that every use of REBOA is harmful nor a neutral result that can be dismissed because the primary credible interval crosses one. The interval permits substantial benefit and substantial harm, but the posterior distribution gives considerably more weight to harm. The 86.9% probability describes the treatment effect under the specified model and prior; it is not an individual patient's probability of dying.¹

Randomisation before the decision pathway unfolded is a major strength. Successful inflation in only 41% of allocated patients does not invalidate the intention-to-treat question. Failed access, improvement during preparation and deterioration before deployment are practical consequences of offering the strategy. Restricting analysis to the nineteen successfully treated patients would select on events occurring after allocation and discard much of the question facing an emergency department. Nevertheless, the trial cannot precisely estimate the biological effect of inflation in an identifiable subgroup known in advance to require and receive it.¹

The eight failed-access cases are especially relevant to implementation. They show that obtaining arterial access in this population was a substantial obstacle despite organised

training. Conversely, the seventeen patients who improved suggest that initial judgement of likely need often changed during resuscitation. These two pathways should not be collapsed into a single label of non-adherence: one represents technical failure, the other a clinical decision that proceeding to inflation was unnecessary. A policy demanding inflation in every allocated patient would have tested a different, less clinically responsive intervention.^{1,3}

Chance imbalance remains important in such a small trial. The REBOA group had lower admission pressure, more patients with extreme hypotension, and a higher median head-injury score. Identical median overall Injury Severity Scores do not guarantee equivalent risks of early exsanguination or death from brain injury. Adjustment did not reverse the direction of the result, but the extensive covariate model was supported by only 43 deaths and produced a wide interval. Missing blood pressures and resuscitation variables required assumptions. These analyses reduce confidence in a simple imbalance explanation without proving that imbalance had no influence.^{1,3}

The Bayesian framework is well suited to a rare, high-risk emergency where a fixed large trial may be infeasible. It also makes the uncertainty explicit. The reduction in the estimated probability of harm from 90.1% at stopping to 86.9% in the completed dataset is expected to be possible as further outcomes arrive; it does not mean the stopping calculation was incorrectly applied. Early termination limits precision, however, and the acknowledged design error and subsequent revisions deserve scrutiny. The principal report does not supply enough detail to reconstruct the exact amendment chronology or fully audit the original versus revised operating characteristics.¹

The timing of the deaths strengthens concern about the immediate resuscitation pathway. Six- and three-hour estimates were more adverse than the 90-day estimate. Yet these are overlapping outcomes involving many of the same deaths, not independent replications, and three-hour mortality was added during the trial. Multiple time points and a small sample caution against elevating the most extreme estimate into the definitive effect size. Their appropriate role is to support concern about early harm while the primary outcome remains the principal estimate.^{1,3}

The longer interval to commencement of definitive haemorrhage control provides a plausible explanation, but does not establish mediation. It was measured only among patients reaching a qualifying procedure, and procedure receipt itself differed between groups. Patients dying early could not contribute a subsequent operative time. The reported nineteen-minute difference between medians is therefore not a randomised estimate of delay for every participant. The restrictive procedure definition also omitted some clinically relevant operations. The data suggest a pathway problem worth

investigating; they cannot determine how many deaths were caused by diversion of time to REBOA.^{1,3}

Similarly, the absence of reported serious adverse device events does not exclude treatment-related harm. A strategy can worsen survival through delay, ischaemia or altered resuscitation without a recognised device malfunction. More early deaths also reduce the opportunity to record later complications or transfusions. Hospital-free days appropriately assign death the worst value, but their adverse result largely reinforces the mortality finding; it should not be interpreted as an isolated effect on survivors' length of stay. Cause-of-death attribution adds clinical context, although eight deaths had an unknown cause and attribution is less objective than all-cause mortality.^{1,3}

The accompanying editorial questioned training, operator experience, devices and generalisability to highly experienced centres. These are reasonable domains of uncertainty, but the assertion that training was not described is qualified by the substantial programme presented in the appendix. What remains unavailable is a sufficiently detailed account of operator-level competence and clinical exposure. Trial recruitment frequency is not necessarily total procedural experience. Equally, the 32-minute arrival-to-inflation interval cannot be relabelled as the time required to perform cannulation. Fair criticism should preserve those distinctions.^{3,4}

Exploratory analyses do not resolve these uncertainties. Removing the first REBOA allocation at each enrolling centre is not a controlled test of training or an operator learning curve, and it changes only one randomised group. Principal-stratum analyses depend on assumptions about treatment receipt under alternative allocations and were extremely imprecise. Some supplementary subgroup counts are internally inconsistent with the stated definitions and main outcomes, preventing reliable use of their numerical estimates. These reporting problems limit those analyses without erasing the independently reported primary counts.³

Applicability is strongest to the severely injured, predominantly blunt-trauma population and emergency-department systems actually studied. Only three patients had penetrating injuries; comparisons between balloon zones, partial versus complete occlusion, or prehospital strategies are not established. The trial does not identify a subgroup with proven benefit. Its practical contribution is to make routine expansion of this resource-intensive intervention difficult to justify on an assumption of improved survival, while leaving specialised alternative pathways as questions requiring direct evaluation.¹

Body of Evidence

The earlier evidence illustrates how choice of comparator can change the apparent effect of REBOA. Joseph and colleagues used US Trauma Quality Improvement Program data from 2015–2016 to match 140 REBOA patients with 280 controls. Mortality was 35.7% versus 18.9%, with more acute kidney injury and lower-extremity amputation among REBOA recipients. Matching included physiology and anatomical injury measures, but the study remained retrospective: the evolving clinical reasons for deploying REBOA could not be randomised away. Its adverse association raised concern, without proving that the procedure caused the difference.⁵

Harfouche and colleagues reported the opposite association from a single high-volume centre. Fifty-seven REBOA patients were matched separately with 114 contemporary and 114 historical controls; patients arriving in cardiac arrest were excluded. In-hospital mortality was 19.3% with REBOA, 35.1% in contemporary controls and 44.7% in historical controls. A specialist endovascular trauma service supported the intervention. This establishes an encouraging association in a particular system, not a transportable causal benefit. Historical comparison remains vulnerable to changes in care over time, while contemporary selection retains uncertainty about why otherwise similar patients received different treatment.⁶

Cralley and colleagues examined a different decision: zone I REBOA versus resuscitative thoracotomy among patients undergoing aortic occlusion in the AORTA registry. Of 991 patients, propensity matching produced 56 pairs. Mortality in those matched groups was 78.6% with REBOA versus 92.9% with thoracotomy. The extremely high mortality and the requirement for aortic occlusion distinguish this population and comparator from a general policy of adding REBOA to standard care. An association favouring one method of aortic occlusion cannot establish that attempting occlusion is preferable to proceeding without it. Matching also cannot eliminate unmeasured differences in treatment selection.⁷

The 2019 joint statement placed system organisation at the centre of safe use. It required suitably trained clinicians and expeditious access to definitive haemorrhage control, while acknowledging inadequate high-grade evidence. These were conditions proposed to reduce risk, rather than proof that adherence delivered a survival advantage. UK-REBOA tested an implemented strategy with training and specialist trauma services, but did not randomise alternative training programmes or models of service organisation. Its result therefore informs the benefit–harm balance without settling the best possible delivery model.^{1,2,3}

The editorial's argument that expertise and procedural details may explain different results is a plausible hypothesis. It should be tested with prospectively defined pathways and explicit measures of access success, occlusion duration and progress to definitive haemostasis. The trial provides no controlled evidence that a higher-volume service would reverse the mortality result. Conversely, its small size and mixed delivery do not warrant a universal claim that every carefully selected REBOA use must be harmful. Both interpretations would exceed the randomised evidence available by the end of 2023.^{1,4}

The evidence hierarchy matters most when the clinical intervention is selected in response to deterioration. Observational harm may reflect preferential treatment of the sickest patients; observational benefit may reflect selection of patients likely to survive long enough for successful deployment or particularly capable centres. UK-REBOA reduces those selection problems by assigning the initial strategy, even though chance imbalance and imprecision remain. Its findings should carry greater weight for that policy decision than contradictory matched-cohort associations. The unresolved task is to identify and test a clearly specified alternative pathway with a credible mechanism for benefit, while preserving rapid definitive haemorrhage control as the central objective.^{1,5,6,7}

Should REBOA routinely be added to emergency-department care for exsanguinating torso trauma?

UK-REBOA found no survival benefit and a concerning signal of early harm. Routine addition of the tested strategy is not supported.

References

1. Jansen JO, Hudson J, Cochran C, MacLennan G, Lendrum R, Sadek S, et al; UK-REBOA Study Group. Emergency Department Resuscitative Endovascular Balloon Occlusion of the Aorta in Trauma Patients With Exsanguinating Hemorrhage: The UK-REBOA Randomized Clinical Trial. *JAMA*. 2023;330(19):1862–1871.
2. Bulger EM, Perina DG, Qasim Z, Beldowicz B, Brenner M, Guyette F, et al. Clinical use of resuscitative endovascular balloon occlusion of the aorta (REBOA) in civilian trauma systems in the USA, 2019: a joint statement from the American College of Surgeons Committee on Trauma, the American College of Emergency Physicians, the National Association of Emergency Medical Services Physicians and the National Association of Emergency Medical Technicians. *Trauma Surg Acute Care Open*. 2019;4(1):e000376.

3. Jansen JO, Hudson J, Cochran C, MacLennan G, Lendrum R, Sadek S, et al. UK-REBOA supplementary online content (Supplement 1). JAMA. 2023.
4. Tisherman SA, Brenner ML. Contemporary Adjuncts to Hemorrhage Control. JAMA. 2023;330(19):1849–1851.
5. Joseph B, Zeeshan M, Sakran JV, Hamidi M, Kulvatunyou N, Khan M, et al. Nationwide Analysis of Resuscitative Endovascular Balloon Occlusion of the Aorta in Civilian Trauma. JAMA Surg. 2019;154(6):500–508.
6. Harfouche MN, Madurska MJ, Elansary N, Abdou H, Lang E, DuBose JJ, et al. Resuscitative endovascular balloon occlusion of the aorta associated with improved survival in hemorrhagic shock. PLoS One. 2022;17(3):e0265778.
7. Cralley AL, Vigneshwar N, Moore EE, Dubose J, Brenner ML, Sauaia A; AAST AORTA Study Group. Zone 1 Endovascular Balloon Occlusion of the Aorta vs Resuscitative Thoracotomy for Patient Resuscitation After Severe Hemorrhagic Shock. JAMA Surg. 2023;158(2):140–150.

RESTRIC

Hayakawa M, Tagami T, Kudo D, Ono K, Aoki M, Endo A, et al. The Restrictive Red Blood Cell Transfusion Strategy for Critically Injured Patients (RESTRIC) trial: a cluster-randomized, crossover, non-inferiority multicenter trial of restrictive transfusion in trauma. J Intensive Care. 2023;11:34.

Introduction

Red-cell transfusion after severe injury must reconcile two aims: preserving oxygen delivery during blood loss and avoiding exposure that provides no additional clinical benefit. A haemoglobin concentration measured early in haemorrhage cannot, by itself, describe the circulation's immediate needs. Clinicians may need to transfuse before a substantial fall is documented, while later transfusions may chiefly determine how high a haemoglobin concentration is maintained. Evidence about stable intensive-care anaemia cannot automatically settle both decisions.¹

The Transfusion Requirements in Critical Care trial, TRICC, established an important basis for restrictive practice in critically ill patients after initial treatment. It compared transfusion thresholds of 7 and 10 g/dL, maintaining haemoglobin at 7–9 and 10–12 g/dL respectively. Its population was euvolaemic when enrolled, and the timing and clinical circumstances differed from emergency management of a patient with potentially uncontrolled traumatic bleeding.²

RESTRIC's investigators identified this gap in the evidence supporting haemoglobin targets in trauma. They sought to apply a transfusion policy immediately on emergency-department arrival, including patients judged at risk of severe bleeding before the eventual transfusion requirement was known. The intervention was deliberately framed as a target range, allowing the clinician to initiate blood according to haemodynamic instability as well as haemoglobin. It was not a rule requiring all patients to wait until haemoglobin fell below a fixed threshold.¹

The hypothesis was that a restrictive target would conserve red cells without an unacceptable loss of survival compared with a liberal target. This makes the trial's non-inferiority objective fundamental: reduced transfusion alone would not be sufficient if clinically important harm could not be excluded. RESTRIC therefore needed both a genuine difference in treatment delivery and enough precision to judge whether any survival disadvantage remained within its prespecified margin.¹

Synopsis

RESTRIC was a pragmatic cluster-randomised, crossover, non-inferiority trial conducted at 22 tertiary emergency medical institutions in Japan. Recruitment occurred from 7 May 2019 to 31 October 2021. Support came partly from the General Insurance Association of Japan and the Marumo Emergency Medical Research Promotion Fund; the funding bodies were reported to have no role in design, data collection or interpretation, or manuscript preparation.¹

Institutions were assigned equally to one of two treatment sequences using a previously prepared random-assignment table. Each centre applied its first strategy for one year, followed by a one-month washout and one year using the other strategy. The allocation therefore concerned a hospital's treatment policy during a period, not an individually concealed randomisation at the bedside. The policy was displayed at each centre, and treatment was unblinded. Written consent for study enrolment was obtained from patients or their representatives as soon as possible after initial care; those declining enrolment subsequently received treatment at their physician's discretion.¹

Patients aged at least 20 years qualified if the treating physician judged that they had severe bleeding capable of causing shock, suspected severe bleeding after arrival, or a potential for major postoperative bleeding during the acute injury phase. Exclusions included cardiac arrest before or at arrival, transfer from another hospital, an initial decision to withdraw active treatment, burns involving at least 15% of body surface area, pregnancy, chronic anaemia with haemoglobin at or below 7 g/dL, and objection to transfusion. Entry did not require a particular admission haemoglobin, shock index or number of red-cell units already administered.¹

The restrictive strategy targeted haemoglobin of 7–9 g/dL and the liberal strategy 10–12 g/dL. Clinicians determined when to start red cells during active bleeding using haemoglobin and haemodynamic instability. Crystalloid or colloid could be given for resuscitation when haemoglobin was sufficiently high for the assigned target. The restrictive policy was explicitly not a strategy of permissive hypotension or hypovolaemia. Treatment continued until seven days after admission, intensive-care discharge, withdrawal of active treatment or death. Plasma, platelets and surgical or radiological haemostasis formed part of care, but the report does not specify a single mandatory component ratio applied in both groups.¹

The primary outcome was survival at 28 days after emergency-department arrival. Patients discharged earlier were contacted by telephone, and those with an unknown day-28 status were excluded from analysis. Secondary outcomes included time to death, cumulative red-cell, plasma and platelet use, organ failure during seven days, selected

complications during 28 days, Glasgow Outcome Scale at hospital discharge, and intensive-care-, ventilator- and catecholamine-free days to day 28. Death within 28 days was assigned zero event-free days. An independent two-member safety board reviewed reports of significant adverse events.¹

The planned non-inferiority margin allowed a three-percentage-point reduction in survival. Investigators assumed 25% mortality in the liberal group, correlation coefficients of 0.05 within clusters and periods, 80% power and a one-sided significance level of 2.5%. Their original calculation produced 170 participants per group, with a target of 400 allowing for cluster-size variation, ineligible inclusions and loss to follow-up. However, the publication acknowledges that this calculation was erroneous: the corrected requirement was 6,214 patients. The error was discovered only after the study had finished.¹

The primary logistic mixed model included treatment and period as fixed effects and site and site-by-period interaction as random effects. The main analysis followed allocation in the full analysis set after exclusions, with a per-protocol sensitivity analysis removing intentional departures from the assigned target. Non-inferiority required the lower confidence limit to remain above the permitted survival disadvantage. The report expressed this on the odds-ratio scale using a boundary of 0.680, described as corresponding to the three-point margin and an adjusted liberal-group survival of 93.4%.¹

The flow diagram records 5,935 patients aged at least 20 years with severe trauma and 1,045 eligible patients. Of these, 543 were in restrictive periods and 502 in liberal periods; 321 and 302 respectively did not consent. Thus, 222 and 200 were enrolled. Five subsequently excluded inclusions and six patients with unavailable day-28 status left 216 restrictive and 195 liberal patients for the main analysis. The six lost to follow-up had left hospital alive but could not be contacted. Six restrictive and one liberal participant were then excluded for intentional treatment deviations, giving per-protocol populations of 210 and 194.¹

Median ages were 61 and 57 years, about two-thirds were men, and approximately nine in ten had blunt trauma. Median Injury Severity Scores were 24.5 and 24.0. Admission systolic pressures were 103.5 and 110.0 mm Hg, and haemoglobin concentrations 12.2 and 11.9 g/dL. Major haemostatic interventions during six hours were recorded in 116/216 restrictive patients and 130/195 liberal patients, 53.7% versus 66.7%; chest interventions occurred in five versus eighteen. The supplementary site table showed highly uneven recruitment, with some institutions contributing analysed patients under only one strategy.^{1,3}

Haemoglobin fell rapidly in both groups. Thereafter, the restrictive group generally remained around 9 g/dL and the liberal group around 10–11 g/dL through the first week. The red-cell curves showed no clear difference in cumulative use during the initial six hours; higher liberal-group use emerged subsequently. More restrictive-group patients avoided red-cell transfusion. Cumulative plasma use was not significantly different. Thus, the achieved contrast involved lower haemoglobin and less red-cell exposure, with limited separation during the earliest resuscitation hours.¹

At 28 days, 199/216 restrictive patients and 178/195 liberal patients were alive, 92.1% versus 91.3%. The adjusted odds ratio for survival was 1.017, with a 95% confidence interval of 0.485–2.131. The lower limit was below the reported non-inferiority boundary of 0.680, so non-inferiority was not demonstrated. There were seventeen deaths in each group. Per-protocol survival was 193/210 versus 177/194, 91.9% versus 91.2%; the supplied per-protocol table did not provide a separate adjusted confidence interval establishing non-inferiority.^{1,3}

Respiratory failure occurred in 23 versus eighteen patients and renal failure in seven versus eleven during the first week. At 28 days, cerebral infarction and pulmonary embolism each occurred in five versus two patients, deep-vein thrombosis in 24 versus seventeen, and sepsis in five versus thirteen. There were no recorded myocardial infarctions or transfusion-related acute lung injury episodes. Median intensive-care-free days were nineteen in each group, ventilator-free days 25.5 versus 24, and catecholamine-free days 28 in each. The report identified no significant differences in these outcomes or discharge Glasgow Outcome Scale. Subgroup interaction tests likewise did not identify convincing variation in the survival effect.^{1,3}

Critique

RESTRIC did not answer its principal non-inferiority question. The limitation is more substantial than a modest recruitment shortfall: the original sample-size calculation was wrong by an order of magnitude. Enrolling slightly more than the intended 400 participants therefore did not deliver the intended precision. The similar observed survival percentages cannot substitute for excluding the prespecified margin, and neither the absence of statistical superiority nor favourable red-cell use establishes acceptable safety.¹

The confidence interval is the clearest expression of what remains unknown. For the survival outcome, odds ratios above one favour restriction and values below one favour the liberal strategy. The lower bound of 0.485 crossed well beyond the reported 0.680 boundary. Clinically important disadvantage therefore remained compatible with the

data. Equally, failing non-inferiority does not establish that the liberal strategy is better. The trial is inconclusive about comparative survival, while informative about the ability to reduce red-cell exposure.¹

The three-percentage-point margin also needs clinical context. It permits three additional deaths per hundred patients in exchange for the proposed advantages of restriction. Whether that trade-off is acceptable is a clinical judgement, not a consequence of statistical convention. Actual mortality was much lower than the assumed 25%, making the same absolute allowance proportionally more consequential. The publication's conversion of the margin to an odds-ratio boundary is not fully reproducible from the rounded survival value printed alongside it; this does not change the conclusion because the observed lower confidence limit is materially below the boundary.¹

The cluster crossover design has a practical rationale. Assigning a policy in advance permits treatment from arrival without delaying resuscitation for individual randomisation, and having centres use both policies reduces confounding by stable differences between hospitals. Its quality should not be judged simply by whether a double-blind individual trial would sound stronger. The relevant questions are whether patient selection differed between known treatment periods, whether practice changed over time, and whether the analysis represented the correlation structure. The mixed model addressed site and period effects, although sparse and uneven recruitment limits the information available to estimate them.^{1,3}

Selection after allocation is a particularly important vulnerability. More than half of eligible patients did not consent, and enrolment occurred after the initial assigned policy had already been applied. The proportions not enrolled were broadly similar between strategies, but that does not establish that the clinical reasons were comparable. The available report does not characterise outcomes for these patients sufficiently to exclude differential selection related to early response or prognosis. The analysed population is consequently narrower than all eligible patients exposed to the hospital policy.¹

The eleven exclusions after enrolment further qualify the intention-to-treat label. Some remove clear eligibility problems, but the flow diagram also includes a patient given another strategy. Excluding a treatment error from the main analysis can weaken preservation of allocation. Four restrictive and two liberal patients had missing day-28 status despite discharge alive. This is a small absolute amount of missingness, but there were only seventeen observed deaths per group. Treating discharge alive as proof of

day-28 survival would be unjustified; excluding those patients instead relies on an assumption that their unavailable outcomes do not materially distort comparison.¹

The intervention's flexibility was appropriate to bleeding physiology, but limits what can be inferred about thresholds. Clinicians could transfuse before haemoglobin reached the target range, and median red-cell use initially overlapped. Later haemoglobin separation shows that the policies altered care, yet the restrictive arm mostly occupied the upper end of its intended range. The trial provides more direct evidence about maintaining haemoglobin near 9 rather than 10–11 g/dL after initial management than about permitting every patient to reach 7 g/dL during uncontrolled haemorrhage. A trigger-only interpretation would misdescribe the tested strategy.¹

Important co-interventions also differed. The liberal group underwent more major haemostatic procedures, especially in the chest. The main prose describes intervention frequencies as not significantly different, but Table 2 reports 53.7% versus 66.7% for major haemostasis, with $P=0.039$. These procedures occurred after assignment, so the difference could reflect injury patterns, selection, clinical decisions or consequences of treatment; its cause cannot be determined from the table. It is not appropriate either to ignore the difference or to adjust it away automatically as though it were a baseline characteristic.¹

The main model adjusted for the trial's design, not a comprehensive set of patient-level prognostic variables. Baseline medians were broadly comparable, but non-significant comparisons do not certify balance in bleeding severity or physiology. The broadly defined eligibility criteria admitted patients at risk of bleeding, including some never transfused, while excluding cardiac arrest and transferred patients. This helped test a practical policy but reduced representation of the most immediately unstable population. The modest mortality and limited early transfusion separation should temper extrapolation to exsanguinating shock.¹

Safety findings are too sparse for categorical reassurance. No myocardial infarctions or acute transfusion-related lung injury episodes were recorded, but uncommon harms cannot be excluded in 411 participants. Numerical directions differed between complications, and testing multiple secondary outcomes without a demonstrated primary non-inferiority result cannot establish safety for each organ system. The report's suggestion that restriction had no ischaemic effect exceeds what these event counts can show. Discharge Glasgow Outcome Scale is also not a substitute for adequately powered, longer-term neurological assessment in severe traumatic brain injury.^{1,3}

Per-protocol analysis is useful in a non-inferiority trial, where contamination may make treatments appear similar. Here, removing only seven patients left comparable survival percentages, but percentages alone are not a margin-based analysis. Intentional deviations also do not describe every haemoglobin measurement outside target or quantify the time patients spent at each concentration. The supplied sensitivity material therefore does not rescue the inconclusive primary test. Likewise, absent subgroup interactions cannot establish equivalent effects across head injury or injury-severity categories, especially when the entire trial has limited precision.^{1,3}

RESTRIC's defensible contribution is consequently specific. A pragmatic lower-target policy changed haemoglobin and reduced red-cell use in the enrolled Japanese trauma population. It did not establish that the resulting survival disadvantage, if any, was smaller than three percentage points. That distinction preserves the useful implementation finding while avoiding an unsupported clinical safety conclusion.¹

Body of Evidence

TRICC randomised 838 euvolaemic critically ill patients with haemoglobin below 9 g/dL within 72 hours of intensive-care admission. Thirty-day mortality was 18.7% with restrictive transfusion and 23.3% with liberal transfusion, $P=0.11$. Its entry conditions are central to comparison with RESTRIC: the trial addressed transfusion after initial treatment, not an emergency-department strategy for suspected ongoing major bleeding. The findings challenged routine maintenance of a higher haemoglobin concentration, but cannot simply be transferred to every phase of trauma resuscitation.²

RESTRIC's authors also highlighted the post hoc trauma analysis of TRICC as an important basis for earlier recommendations. As they described it, those patients had already reached intensive care, and active bleeding had been excluded. This is a subgroup of an existing trial, not a second independent emergency-resuscitation experiment. The distinction helps explain why a trauma-specific trial was needed, even when the nominal restrictive and liberal haemoglobin ranges were familiar. It also prevents double-counting the parent trial and its trauma subgroup as separate randomised populations.^{1,2}

Robertson and colleagues studied a related concern in 200 patients with closed traumatic brain injury, randomising transfusion thresholds of 7 or 10 g/dL within a factorial erythropoietin trial. Favourable six-month neurological outcome occurred in 37/87 evaluable patients at the lower threshold and 31/94 at the higher threshold, 42.5% versus 33.0%, $P=0.28$. Thromboembolic events were more frequent in the higher-threshold group. Patients with life-threatening systemic injuries were excluded,

so this does not directly answer how to transfuse a patient with combined major haemorrhage and brain injury. It also does not convert RESTRIC's small head-injury subgroup into definitive neurological safety evidence.⁴

PROPPR examined component composition during active trauma resuscitation. Among 680 patients predicted to require massive transfusion, plasma, platelets and red cells in a 1:1:1 ratio were compared with 1:1:2. All-cause mortality did not differ significantly at 24 hours or 30 days, although more patients achieved haemostasis and fewer died from exsanguination within 24 hours with 1:1:1. Median red-cell use over 24 hours was nine units in both groups. A ratio strategy therefore cannot be equated with a liberal haemoglobin target: the trials manipulated different aspects of transfusion, and balanced component delivery does not itself define the haemoglobin concentration that should later be maintained.⁵

The sixth European trauma-bleeding guideline, published in March 2023 before RESTRIC's report, recommended a haemoglobin target of 70–90 g/L when red-cell transfusion was necessary, graded 1C. Its rationale explicitly acknowledged the absence of acute bleeding trauma trials in the broader threshold evidence and warned against basing transfusion decisions on haemoglobin alone during rapidly evolving haemorrhagic shock. This places the recommendation and RESTRIC in their proper historical sequence: the guideline supplied a pragmatic target under uncertainty; RESTRIC subsequently provided direct policy data without conclusively establishing non-inferiority.⁶

Together, these studies separate three questions that can otherwise become conflated: emergency replacement during blood loss, the balance of components supporting haemostasis, and the haemoglobin target maintained after initial resuscitation. RESTRIC brought the target policy forward to arrival, but its observed transfusion contrast became clearer later. Its results should therefore be interpreted alongside the earlier evidence rather than used to replace physiological judgement with a single number. By the end of 2023, the trial supported the feasibility of reducing red-cell use in this setting, while leaving the survival trade-off insufficiently resolved. A definitive answer would require a correctly sized study with careful handling of recruitment after cluster allocation, early treatment separation and missing outcomes.^{1,2,5,6}

Is a haemoglobin target of 7–9 g/dL non-inferior to 10–12 g/dL from arrival after major trauma?

RESTRIC reduced red-cell use but did not establish non-inferiority for 28-day survival. Similar observed survival does not exclude clinically important harm.

References

1. Hayakawa M, Tagami T, Kudo D, Ono K, Aoki M, Endo A, et al. The Restrictive Red Blood Cell Transfusion Strategy for Critically Injured Patients (RESTRIC) trial: a cluster-randomized, crossover, non-inferiority multicenter trial of restrictive transfusion in trauma. *J Intensive Care*. 2023;11:34.
2. Hébert PC, Wells G, Blajchman MA, Marshall J, Martin C, Pagliarello G, et al; Transfusion Requirements in Critical Care Investigators, Canadian Critical Care Trials Group. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. *N Engl J Med*. 1999;340(6):409–417.
3. Hayakawa M, Tagami T, Kudo D, Ono K, Aoki M, Endo A, et al. RESTRIC supplementary results: Additional files 3–7. *J Intensive Care*. 2023.
4. Robertson CS, Hannay HJ, Yamal JM, 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. Holcomb JB, Tilley BC, Baraniuk S, Fox EE, Wade CE, Podbielski JM, et al; PROPPR Study Group. Transfusion of Plasma, Platelets, and Red Blood Cells in a 1:1:1 vs a 1:1:2 Ratio and Mortality in Patients With Severe Trauma: The PROPPR Randomized Clinical Trial. *JAMA*. 2015;313(5):471–482.
6. Rossaint R, Afshari A, Bouillon B, Cerny V, Cimpoesu D, Curry N, et al. The European guideline on management of major bleeding and coagulopathy following trauma: sixth edition. *Crit Care*. 2023;27:80.

PATCH-TRAUMA

Gruen RL, Mitra B, Bernard SA, McArthur CJ, Burns B, Gantner DC, et al.; PATCH-Trauma Investigators and the ANZICS Clinical Trials Group. Prehospital Tranexamic Acid for Severe Trauma. N Engl J Med. 2023;389(2):127–136.

Introduction

Tranexamic acid offers an unusually simple intervention for a complex emergency. By inhibiting fibrinolysis, it may preserve clot formation during the interval between major injury and definitive haemorrhage control. Its practical attraction is greatest before hospital arrival, when bleeding continues but surgery, interventional radiology and comprehensive laboratory assessment remain unavailable. The relevant question is whether earlier antifibrinolytic treatment improves outcomes when added to an organised trauma system already providing rapid transport, blood products and definitive treatment.¹

CRASH-2 established a reduction in short-term mortality among patients with, or at risk of, significant traumatic bleeding. Its subsequent timing analysis suggested that benefit depended on early administration. However, short-term survival does not describe the eventual consequences of major injury, and the incremental benefit of tranexamic acid in advanced trauma systems remained uncertain. Differences in patient selection, haemorrhage management and access to critical care could affect both the opportunity for benefit and the balance of complications.^{2,3}

Functional recovery is especially important when an intervention might prevent early haemorrhagic death without reversing severe brain injury or other disabling trauma. A treatment can improve survival without increasing the proportion of patients who regain independence. Conversely, assessing recovery at a fixed time may miss later improvement. These are distinct clinical questions, requiring outcomes that extend beyond hospital discharge and mortality alone.

PATCH-TRAUMA therefore tested prehospital tranexamic acid followed by an in-hospital infusion in severely injured adults at high risk of trauma-induced coagulopathy. Its hypothesis was that treatment within three hours of injury would increase survival with a favourable functional outcome at six months. The trial deliberately placed quality of survival, rather than early mortality alone, at the centre of its assessment.¹

Synopsis

PATCH-TRAUMA was an international, double-blind, placebo-controlled superiority trial involving 15 emergency medical services and 21 hospitals in Australia, New Zealand and Germany. Recruitment ran from 28 July 2014 to 28 September 2021. Funding came from public and charitable bodies in the participating countries, including the Australian National Health and Medical Research Council; there was no commercial support. Emergency enrolment proceeded under a waiver of prospective consent, followed by notification and consent for continued participation and data collection when feasible.¹

Eligible patients were adults aged at least 18 years with suspected severe traumatic injury, transported to a participating trauma centre, in whom treatment could begin before hospital admission and within three hours of injury. A Coagulopathy of Severe Trauma (COAST) score of at least 3 identified increased risk. This seven-point score incorporates vehicle entrapment, hypotension, hypothermia, suspected pneumothorax and suspected intra-abdominal or pelvic injury, with additional points for more severe hypotension or hypothermia. Known or suspected pregnancy and residence in an aged-care facility were exclusions. Eligibility did not require laboratory confirmation of coagulopathy or hyperfibrinolysis.

Allocation was 1:1 using computer-generated random numbers, stratified by national or Australian state jurisdiction and initial Glasgow Coma Scale score below 9 versus at least 9. Independent packaging companies supplied consecutively numbered, opaque, sealed packs containing two identical ampoules. Participants, treating clinicians and follow-up assessors remained blinded. Each active ampoule contained 1 g of tranexamic acid; placebo contained 0.9% sodium chloride.

The first dose was given intravenously over ten minutes at the scene or during transport. After hospital arrival, the second ampoule was added to one litre of saline and infused over eight hours. Both groups otherwise received usual prehospital, hospital and subsequent care. Treatment could stop for suspected allergy, recognition of an exclusion, transition to palliation or refusal of further participation. Inpatients underwent screening for lower-limb deep venous thrombosis by Doppler ultrasonography around day 7. The comparison therefore concerned an early drug regimen added to trauma care, without imposing a common transfusion or operative algorithm.¹

The primary endpoint was survival with a favourable functional outcome six months after injury, defined as a Glasgow Outcome Scale–Extended (GOS-E) score of 5–8. Scores 1–4 comprise death, vegetative state and severe disability; scores 5–8 range from moderate disability to good recovery. Trained, blinded assessors used structured telephone

interviews with patients or appropriate proxies. Mortality at 24 hours, 28 days and six months was secondary. Vascular occlusive events and sepsis were recorded until death, discharge or 28 days, whichever occurred first. Other outcomes included blood-product use and ventilator-free days, assigning zero to patients dying within 28 days. The appendix lists quality-of-life assessments that were not presented in this report.^{1,4}

The initial sample of 1,184 was calculated to provide 90% power to detect a nine-percentage-point difference in favourable survival at a two-sided alpha of 0.05. The principal report does not state the assumed control proportion. In February 2020, the target increased to 1,316 because observed loss to follow-up was approximately 10%. Recruitment ended at 1,310 when trial agents expired. Independent monitoring included planned reviews after 296 and 592 patients completed 28-day follow-up. The report dates public posting of the analysis plan to 19 August 2022, before unblinding.

The primary analysis used available outcomes in the intention-to-treat population, excluding unknown assignments and withdrawals of permission to use any data. Log-binomial regression estimated the risk ratio. Supplementary analyses adjusted for stratification variables, restricted assessment to four to eight months, or used 30 multiply imputed datasets. Per-protocol analyses excluded eligibility and treatment deviations. Secondary comparisons were exploratory, without correction for multiplicity; their confidence intervals were not intended to substitute for confirmatory hypothesis tests.^{1,4}

Of 1,310 recruits, three had unknown assignments because trial packs were lost. There were 661 known allocations to tranexamic acid and 646 to placebo; seven subsequent withdrawals left 657 and 643 in the intention-to-treat population. No screening denominator or reasons for non-enrolment were collected. Primary outcomes were available for 572 and 559 patients: 50 declined follow-up and 119 could not be contacted. Mean age was approximately 44 years, around 70% were male and over 90% had blunt trauma. Median Injury Severity Score was 29 in both groups. More than 70% had an initial systolic pressure below 90 mmHg, and approximately 36% had already received red cells before randomisation. Laboratory evidence of early coagulopathy was present in 24.1%.¹

The first dose was administered within one hour in 32.6% versus 27.5%, and between one and two hours in 45.3% versus 51.3%. Only the first study dose was received by 140/657 (21.3%) versus 163/643 (25.3%). Open-label tranexamic acid was given in hospital to 104/657 (15.8%) versus 106/643 (16.5%). Any deviation occurred in 32.7% versus 37.0%, leaving 442 versus 405 patients in the per-protocol population. These categories overlap and cannot be added.⁴

Favourable six-month survival occurred in 307/572 (53.7%) versus 299/559 (53.5%); risk ratio 1.00, 95% confidence interval 0.90–1.12; $P=0.95$. The reported absolute difference was 0.2 percentage points, with a 95% confidence interval from –5.6 to 6.0. Imputation estimated 56.1% in each group, with a risk ratio of 1.00 (0.90–1.11). Adjustment, restriction of interview timing and per-protocol analysis gave similar overall conclusions. No prespecified subgroup showed evidence of treatment-effect heterogeneity.^{1,4}

Mortality was 64/657 (9.7%) versus 90/640 (14.1%) at 24 hours, risk ratio 0.69 (0.51–0.94); 113/653 (17.3%) versus 139/637 (21.8%) at 28 days, risk ratio 0.79 (0.63–0.99); and 123/648 (19.0%) versus 144/629 (22.9%) at six months, risk ratio 0.83 (0.67–1.03). Death attributed to bleeding occurred in 36 versus 52 patients, with a cause-specific hazard ratio of 0.66 (0.43–1.01). In the complete-case GOS-E distribution, lower severe disability occurred in 15.7% versus 12.0%, and upper severe disability in 9.1% versus 8.4%.¹

Any vascular occlusive event occurred in 155/657 (23.6%) versus 126/641 (19.7%), risk ratio 1.20 (0.97–1.48). Deep venous thrombosis occurred in 15.2% versus 12.5%, and pulmonary embolism in 6.5% versus 7.0%. Accounting for competing death produced an estimated risk ratio of 1.13 (0.91–1.41) for any vascular occlusive event. Red-cell transfusion within 24 hours of hospitalisation occurred in 73.1% versus 75.5%; among recipients with quantity data, median use was four versus five bags. Median ventilator-free days were 23 versus 22, difference 1.00 days (–0.55 to 2.55). There was no clear separation in most coagulation measurements.^{1,4}

Critique

PATCH-TRAUMA does not support a large improvement in favourable six-month survival from this tranexamic acid regimen. It also does not establish that early treatment is ineffective. The primary comparison and mortality comparisons answer different questions. Combining death and severe disability into one unfavourable category makes the endpoint clinically meaningful for independence, but insensitive to a shift from death to survival with substantial disability. The observed distribution makes that distinction central, rather than a technical qualification.¹

The near-identical primary proportions are persuasive against the nine-percentage-point improvement used for planning. However, the confidence interval still includes an absolute benefit of six percentage points and harm of a similar magnitude. These would matter clinically. A non-significant superiority result cannot establish equivalence, and the estimated effect applies to a dichotomised scale at one time point. Changes within

the favourable or unfavourable categories do not alter the primary endpoint, even when meaningful to patients and families.

The early mortality findings deserve attention because they are directionally consistent, temporally plausible and aligned with previous randomised evidence. Nevertheless, they were secondary outcomes among numerous comparisons, and the investigators explicitly treated their intervals as exploratory. The six-month mortality interval includes no difference. Adjustment for stratification variables yielded a more favourable six-month estimate, but selecting that result because its interval excludes the null would change the evidential hierarchy. Mortality benefit is supported as a possibility; it is not a replacement positive primary endpoint.^{1,4}

Nor can the excess of severe disability establish that tranexamic acid caused disability in people who would otherwise have recovered well. Randomisation compares groups, not the unobservable alternative outcomes of individual survivors. Preventing death among patients with major irreversible injuries could produce the same distribution. Conversely, one cannot assume that additional survivors would later recover independence. Six months is a substantial follow-up period, but it does not define lifetime recovery, and the absent quality-of-life results limit assessment of how patients valued the states captured by GOS-E.

Missing primary outcomes affect that interpretation. Thirteen per cent of the intention-to-treat population lacked a six-month functional assessment, whereas vital status was known for substantially more patients. The functional endpoint therefore uses a different denominator from the mortality endpoint. Comparing death percentages in the GOS-E figure directly with the overall mortality percentages would be misleading. Balanced missingness between arms is reassuring but does not establish that missing outcomes have the same distribution. Imputation preserved the null primary result, conditional on its missing-at-random assumptions; it cannot recover unmeasured differences among people who declined or could not be reached.^{1,4}

Assessment procedures otherwise protected the endpoint well. Blinded, trained interviewers used a structured instrument, with proxy interviews permitted. This is stronger than an unblinded clinical impression of recovery. However, the four-to-eight-month sensitivity analysis excluded 99 assessments outside that window. Its agreement with the primary analysis is useful evidence that these timing outliers did not determine the result. Telephone ascertainment and the dichotomy still cannot describe the full burden of cognitive, physical and social disability.

Treatment delivery also narrows the inference. Almost all patients received an initial study dose, but approximately a quarter did not receive the intended second dose, while one in six received open-label tranexamic acid. Placebo-arm exposure reduces the drug contrast, and extra active treatment in the intervention arm complicates the intended dose comparison. The intention-to-treat estimate remains the appropriate comparison of allocation to these strategies under the trial's conditions. It is less directly informative about perfect adherence to a two-dose regimen with no rescue exposure.⁴

Agreement with the per-protocol analysis does not remove this limitation. Its exclusions depend on events after randomisation, including receiving open-label treatment and completing intended dosing. The supplementary flow diagram specifically describes first-dose-only exclusions among patients surviving beyond eight hours. Clinical severity, early death and treatment decisions can therefore influence membership of the analysed subgroup. Adjustment for measured injury characteristics cannot guarantee restoration of the original randomised comparison. The particularly low early mortality within the per-protocol groups illustrates why their survival findings should not be treated as independent confirmation.

Selection by COAST was practical for paramedics but was not biological confirmation that fibrinolysis was excessive or responsive to treatment. Most patients lacked the reported laboratory definition of early coagulopathy at hospital arrival. This could dilute benefit in a responsive subgroup, although that explanation remains hypothetical: arrival measurements follow treatment and transport, and routine coagulation tests do not isolate fibrinolysis. The trial neither identifies a validated laboratory selection rule nor justifies delaying early treatment while seeking one.^{1,4}

Safety is less settled than the phrase “no significant difference” implies. The point estimate for any vascular occlusive event favoured placebo and its interval permits a substantial increase. Systematic lower-limb screening strengthens ascertainment relative to studies relying solely on clinically apparent thrombosis. More survivors may also remain at risk of thrombosis, which makes the competing-risk analysis important. That analysis attenuated the estimate but did not exclude harm. Similar counts of other reported adverse events do not establish absence of thrombotic risk, particularly because death, vascular occlusion and sepsis were tabulated separately.⁴

Finally, applicability is strongest to severely injured, predominantly male patients with blunt trauma in organised Australasian systems. Germany contributed only eight participants. The absence of a screening log limits understanding of who was missed during difficult prehospital enrolment. Penetrating injury, burns, pregnancy and frail institutionalised populations are poorly represented or excluded. The trial is a

substantial advance in functional-outcome evidence, but it supports a qualified conclusion: early survival may improve without a demonstrated increase in favourable six-month recovery, and neither a responsive subgroup nor a risk-free regimen has been established.

Body of Evidence

CRASH-2 provides the principal foundation for tranexamic acid in bleeding trauma. It randomised 20,211 adults with, or at risk of, significant bleeding within eight hours of injury to 1 g over ten minutes followed by 1 g over eight hours, or placebo. Among patients analysed, 28-day all-cause mortality was 1,463/10,060 (14.5%) versus 1,613/10,067 (16.0%), risk ratio 0.91 (0.85–0.97). Death attributed to bleeding was also reduced. Its scale and blinded comparison support a survival effect, but it did not answer PATCH-TRAUMA's question about favourable functional survival six months later. These outcomes should be regarded as complementary rather than interchangeable.²

The subsequent CRASH-2 timing analysis used the same participants and must not be counted as independent replication. For treatment within one hour, bleeding death was 5.3% versus 7.7%, risk ratio 0.68 (0.57–0.82); between one and three hours it was 4.8% versus 6.1%, risk ratio 0.79 (0.64–0.97). The treatment-by-time interaction supported urgency, although timing itself was not randomised. This analysis explains the rationale for moving treatment into the prehospital interval. PATCH-TRAUMA tested implementation within that early window rather than comparing early and delayed administration directly.³

CRASH-3 extended the question to traumatic brain injury without major extracranial bleeding. It randomised 12,737 patients, of whom 9,202 were treated within three hours. In that early-treatment population, head injury-related death within 28 days was 18.5% versus 19.8%, risk ratio 0.94 (0.86–1.02). A prespecified sensitivity analysis excluding a Glasgow Coma Scale score of 3 or bilateral unreactive pupils yielded a risk ratio of 0.89 (0.80–1.00). Results differed by injury severity, favouring treatment in mild-to-moderate injury. This supports distinguishing salvageable intracranial bleeding from overwhelming injury, but does not validate a retrospective anatomical subgroup as a prehospital treatment rule in PATCH-TRAUMA.⁵

The North American prehospital traumatic brain injury trial provides a closer functional-outcome comparison. Its analysis included 966 patients receiving study drug, with suspected moderate or severe brain injury and systolic pressure at least 90 mmHg. Two active regimens, either a 1-g bolus plus 1-g infusion or a 2-g bolus, were compared with placebo. Favourable six-month GOS-E outcome occurred in 65% of the combined

active groups versus 62% with placebo. The adjusted difference was 3.5 percentage points, with a one-sided 90% lower confidence limit of -0.9; the benefit test gave $P=0.16$. This did not demonstrate improved neurological recovery. Its population was selected for brain injury without shock, so it cannot directly settle the haemorrhagic-shock question, but it reinforces the need to measure recovery beyond early survival.⁶

STAAMP randomised 927 patients during transport to four US trauma centres; 903 remained in its intention-to-treat cohort. Entry required prehospital hypotension or tachycardia within two hours, and active patients received 1 g before hospital arrival followed by one of three in-hospital dosing strategies. Thirty-day mortality was 8.1% versus 9.9%, an imputed difference of -1.8 percentage points (-5.6 to 1.9; $P=0.17$). The cohort had a median Injury Severity Score of 12, substantially lower than PATCH-TRAUMA's 29. Post hoc findings favouring treatment within one hour or in severe shock are hypothesis-generating. They cannot override the overall uncertain estimate, although differences in severity and dose help explain why the trials are not identical tests.⁷

By the end of 2023, the evidence therefore supported early tranexamic acid as a potentially life-saving component of bleeding-trauma care while leaving important uncertainty about functional benefit. PATCH-TRAUMA adds a rigorous assessment of that latter question in advanced systems. Its neutral primary result should prevent claims of established improvement in six-month independence; its mortality findings should also prevent dismissal of the earlier survival evidence. The remaining task is to clarify who benefits, how best to deliver treatment without dilution or excess dosing, and whether additional survival translates into outcomes patients would value over longer follow-up.¹

Should prehospital tranexamic acid be used in adults with severe trauma at high risk of coagulopathy?

Consider early treatment for suspected major bleeding. PATCH-TRAUMA supports possible early survival benefit, but does not establish improved favourable recovery at six months.

References

1. Gruen RL, Mitra B, Bernard SA, McArthur CJ, Burns B, Gantner DC, et al.; PATCH-Trauma Investigators and the ANZICS Clinical Trials Group. Prehospital Tranexamic Acid for Severe Trauma. *N Engl J Med*. 2023;389(2):127–136.

2. CRASH-2 trial collaborators. Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): a randomised, placebo-controlled trial. *Lancet*. 2010;376(9734):23–32.
3. Roberts I, Shakur H, Afolabi A, Brohi K, Coats T, Dewan Y, et al.; CRASH-2 collaborators. The importance of early treatment with tranexamic acid in bleeding trauma patients: an exploratory analysis of the CRASH-2 randomised controlled trial. *Lancet*. 2011;377(9771):1096–1101.e2.
4. PATCH-Trauma Investigators and the ANZICS Clinical Trials Group. Supplementary Appendix to: Prehospital Tranexamic Acid for Severe Trauma. *N Engl J Med*. 2023;389:127–136.
5. CRASH-3 trial collaborators. Effects of tranexamic acid on death, disability, vascular occlusive events and other morbidities in patients with acute traumatic brain injury (CRASH-3): a randomised, placebo-controlled trial. *Lancet*. 2019;394(10210):1713–1723.
6. Rowell SE, Meier EN, McKnight B, Kannas D, May S, Sheehan K, et al. Effect of Out-of-Hospital Tranexamic Acid vs Placebo on 6-Month Functional Neurologic Outcomes in Patients With Moderate or Severe Traumatic Brain Injury. *JAMA*. 2020;324(10):961–974.
7. Guyette FX, Brown JB, Zenati MS, Early-Young BJ, Adams PW, Eastridge BJ, et al.; STAAMP Study Group. Tranexamic Acid During Prehospital Transport in Patients at Risk for Hemorrhage After Injury: A Double-blind, Placebo-Controlled, Randomized Clinical Trial. *JAMA Surg*. 2021;156(1):11–20.

PROCOAG

Bouzat P, Charbit J, Abback PS, Huet-Garrigue D, Delhayé N, Leone M, et al.; PROCOAG Study Group. Efficacy and Safety of Early Administration of 4-Factor Prothrombin Complex Concentrate in Patients With Trauma at Risk of Massive Transfusion: The PROCOAG Randomized Clinical Trial. JAMA. 2023;329(16):1367–1375.

Introduction

Severe traumatic haemorrhage requires rapid control of bleeding and simultaneous restoration of circulating volume, oxygen carriage and haemostasis. Plasma, platelets and red blood cells address different components of this problem. Plasma delivers coagulation factors, but obtaining and administering sufficient volumes takes time. A concentrated preparation that can be given promptly could therefore appear attractive when bleeding is brisk and conventional tests suggest impaired coagulation. Whether this approach improves outcomes depends on what physiological deficit is actually present and what other haemostatic treatment the patient receives.

Four-factor prothrombin complex concentrate (4F-PCC) contains factors II, VII, IX and X, together with proteins C and S. Its ability to correct an abnormal prothrombin time does not establish that it improves haemostasis in trauma. Traumatic coagulopathy is heterogeneous, and a prolonged laboratory clotting time need not indicate inadequate thrombin generation. Adding procoagulant factors to a patient who already generates sufficient thrombin could expose that patient to thrombosis without stopping bleeding. The proposed intervention consequently needs evaluation as a clinical strategy, rather than validation through laboratory correction alone.¹

The relevant decision also changes with the intended use of the concentrate. Empirical supplementation of plasma-based resuscitation is different from replacing plasma, correcting an identified factor deficiency or reversing preinjury anticoagulation. Benefits observed with an entire factor-concentrate algorithm cannot automatically be attributed to PCC itself. Earlier observational comparisons encouraged its use, but treatment selection, differences in bleeding severity and other elements of resuscitation could explain some of their apparent benefits.

PROCOAG tested a specific question: would routinely adding an early dose of 4F-PCC to conventional haemostatic resuscitation reduce blood-product consumption in severely injured patients considered at risk of massive transfusion? Its placebo-controlled design

also offered an opportunity to assess the competing concern that empirical factor supplementation might increase thromboembolic complications.

Synopsis

PROCOAG was a double-blind, randomised superiority trial in 12 French academic level I trauma centres. Recruitment ran from 29 December 2017 to 4 August 2021, with follow-up ending on 31 August. The French Ministry of Health funded the study, and LFB supplied Kanokad without charge. The publication states that the funders had no role in its conduct, analysis or publication. Emergency inclusion was permitted, with subsequent consent from the patient or appropriate representative.¹

Adults admitted directly from the scene with the highest level of trauma-team activation were eligible when considered at risk of massive transfusion. The operational assessment combined early red-cell transfusion and an Assessment of Blood Consumption score of at least two, or the treating physician's assessment of risk. Massive transfusion itself was defined as at least three red-cell units during the first hour or ten during the first 24 hours. A prolonged prothrombin time was not required for entry. Important exclusions included traumatic cardiac arrest before randomisation, injuries judged unsurvivable within an hour, secondary transfer, preinjury anticoagulant treatment, pregnancy and known contraindications to PCC. This was therefore a trial of empirical treatment in a selected bleeding-risk population, not anticoagulant reversal.

Allocation was 1:1, stratified by centre, using randomly varying blocks of two to six. An independent statistician generated the sequence. Pharmacy arrangements and numbered sealed envelopes supported allocation concealment. An unblinded nurse outside the clinical team prepared an opaque syringe; patients, treating clinicians and data analysts remained blinded. Randomisation was intended within an hour of admission.

The intervention was intravenous Kanokad at 25 IU of factor IX per kilogram, equivalent to 1 mL/kg, administered at 120 mL/hour as soon as possible after admission. Control patients received the same volume of saline. Both groups otherwise received restricted crystalloid administration, early red cells and fresh frozen plasma at a red-cell:plasma ratio between 1:1 and 2:1, and rapid anatomical haemostasis. Tranexamic acid was recommended within three hours of injury as a 1-g loading dose followed by 1 g over eight hours. Fibrinogen treatment was indicated below 1.5 g/L or with a corresponding viscoelastic deficit, and the platelet target was above $50 \times 10^9/L$. PCC was an additional treatment within this strategy; plasma was not withheld from the intervention group.

The primary endpoint was the total number of red-cell, plasma and platelet units transfused during the first 24 hours from trauma-bay arrival. A lower total favoured PCC. Secondary outcomes included individual product consumption, time to a prothrombin-time ratio below 1.5, anatomical haemostasis, mortality at 24 hours and 28 days, and measures of hospital, intensive-care and ventilator use. Functional outcome at discharge was assessed in patients with relevant head injury. Thromboembolic events were recorded through day 28. Investigation followed clinical suspicion, with ultrasound or contrast-enhanced computed tomography confirmation; systematic venous screening was left to clinicians rather than mandated for every participant.

The sample-size calculation assumed mean consumption of 12 units, with a standard deviation of ten, and sought a three-unit, or 25%, reduction. A log-normal distribution and a Mann–Whitney probability of superiority of 0.59 informed the calculation. A total of 324 patients provided 80% power at a two-sided significance level of 0.05. Additional allocation envelopes allowed for erroneous activations and losses. There was no interim analysis. The primary analysis included randomised patients except those withdrawing consent, with no missing-data imputation. Early deaths contributed the products actually received. Continuous outcomes were compared using rank tests, while the tabulated between-group differences were mean differences with bootstrap confidence intervals. These statistics do not estimate the same feature of the outcome distribution.

Of 4,313 patients screened, 350 were initially considered eligible and 327 were randomised. Consent withdrawal removed one of 165 PCC recipients and two of 162 controls, leaving 164 and 160 patients in the analysis. Approximately 72% were men, the typical age was about 39 years and the median Injury Severity Score was 36. Around four fifths had blunt trauma. Admission systolic pressure was approximately 90 mmHg and lactate approximately 4.6 mmol/L. Among those with available tests, roughly two thirds had a prothrombin-time ratio above 1.2, but only a quarter exceeded 1.5. Substantial bleeding risk did not therefore mean that every participant had severe laboratory coagulopathy.

Median injury-to-arrival times were 105 and 100 minutes. Treatment began a median of 35 minutes after arrival with PCC and 30 minutes with placebo. These separate medians cannot be added to obtain a median injury-to-treatment time. Tranexamic acid was given to 76% and 86%, and fibrinogen to 86% and 81%, respectively. About 70% in each group underwent a haemostatic procedure. Plasma was first administered a median of 73 versus 91 minutes after arrival among patients with recorded timings. The per-protocol population included 308 patients, indicating generally successful delivery within the intended early hospital window.

The primary outcome was a median of 12 units, interquartile range 5–19, with PCC and 11, interquartile range 6–19, with placebo. The reported mean difference was 0.2 units, with a 95% confidence interval from –2.99 to 3.33, and the rank-test P value was 0.72. Red-cell, plasma and platelet requirements were also similar. There was no demonstrated acceleration of anatomical haemostasis. The competing-risk comparison of reaching a prothrombin-time ratio below 1.5 gave a subdistribution hazard ratio of 1.08, 95% confidence interval 0.92–1.28. Mortality at 24 hours was 18/164 versus 20/160. At 28 days, 26 versus 30 deaths were recorded. The SAP specified Kaplan–Meier mortality estimation, so the published percentages should not automatically be interpreted as simple proportions of the randomised groups.²

Thromboembolic events affected 56/161 PCC recipients and 37/157 controls, approximately 35% versus 24%. The absolute difference was 11 percentage points, 95% confidence interval 1–21, and the relative risk 1.48, 95% confidence interval 1.04–2.10; P=0.03. There were 63 versus 46 events, since some patients experienced more than one. Deep-vein thrombosis and pulmonary embolism accounted for most events. Per-protocol and massive-transfusion subgroup analyses did not show reduced product consumption.³

Critique

PROCOAG provides a direct randomised test of routine PCC supplementation during plasma-based trauma resuscitation. The absence of a transfusion benefit, accompanied by more patients experiencing thromboembolic events, gives little support to that strategy. The conclusion should remain attached to the intervention actually tested: one empirical 25-IU/kg dose added to an existing haemostatic programme. It does not establish the value of a different dose, a replacement strategy, laboratory-selected rescue treatment or anticoagulant reversal, which was specifically excluded.

The trial's concealment and blinding are particularly valuable because transfusion is partly a clinician-mediated outcome. Knowledge that PCC had been administered could otherwise influence the decision to request plasma or tolerate continuing bleeding. Matched-volume placebo and separation of drug preparation from clinical care reduce this risk. Multicentre recruitment and the inclusion of both blunt and penetrating injury improve the practical range of the evidence. Almost complete primary-outcome ascertainment and the inclusion of products given before early death further strengthen the immediate comparison.¹

Nevertheless, a count of transfused units is not a direct measure of recovery or haemostatic success. One unit of red cells, plasma and platelets contributes equally to the numerical sum despite different volumes, composition and indications. PCC and

fibrinogen are not included in that product count. Even a favourable result would therefore require interpretation alongside mortality, bleeding control, total treatment exposure and complications. The intervention could shift treatment from counted to uncounted products without delivering a corresponding patient benefit. Here, the absence of a reduction in the counted products means that this potential advantage did not emerge.

Early death also truncates the opportunity to receive transfusion. A patient dying rapidly can have a low product count for an adverse reason. The investigators compared observation times and retained actual consumption, but similar observation-time distributions do not eliminate the competing effect of death. Reassuringly, early mortality was reasonably similar, making a large imbalance in this mechanism less likely. The endpoint still should not be described as an uncomplicated measure of blood saved. Equally, the non-significant primary result does not prove equivalence: the confidence interval around the mean difference extends to approximately three fewer or three more units, close to the difference sought in the sample-size calculation.

The distinction between the statistical summaries also matters. The displayed medians differ by one unit, whereas the reported mean difference is 0.2 units. The P value comes from a rank-based comparison. Substituting one of these estimates for another would change the analysis being described. The publication supports an absence of demonstrated improvement across the distribution, with considerable uncertainty around the average treatment difference; it does not justify reporting a precise one-unit treatment penalty.

Biological selection is central to the applicability of the result. Only about a quarter of tested patients had a prothrombin-time ratio above 1.5. PCC was given before identifying a specific factor deficit, and viscoelastic measurements were available in a minority. This may have exposed some patients with adequate thrombin generation to additional factors. That explanation is plausible, but the trial did not directly establish the mechanism of thrombosis or identify a validated laboratory threshold separating benefit from harm. Post hoc coagulation subgroups cannot supply such a rule.

The safety signal is clinically relevant because there was no compensating benefit in the primary endpoint. Its absolute estimate is also uncertain: the confidence interval spans an increase of one to 21 percentage points. Events of different seriousness were combined, and the study was not designed to establish a precise effect on each individual complication. Detection depended on clinical investigation rather than uniform screening. Blinding helps protect against investigation driven by treatment knowledge, but does not guarantee identical scanning or equal sensitivity in every

patient. The observed event rates describe detected thrombosis under this surveillance system, rather than the complete biological incidence.

Several features complicate comparisons of secondary outcomes. Tranexamic-acid use was lower in the PCC group, while recorded plasma delivery occurred earlier. Fibrinogen was commonly used in both groups. These differences could influence haemostasis, but they arose within the randomised care pathway and should not be treated as if simple post hoc adjustment would recreate a cleaner trial. No convincing separation in transfusion appeared despite high overall protocol adherence. Excluding patients who failed the early-treatment requirement also weakens randomisation, so the per-protocol result is supportive rather than independently decisive.

The subgroup requiring massive transfusion is even more difficult to interpret causally. Membership depends on product administration after randomisation, which the intervention was intended to change. Restricting analysis to that group can select different types of patients in the two arms. Its null result offers no strong evidence for a hidden responsive subgroup, but neither does it establish an unbiased treatment effect among all patients destined to bleed massively. Selecting such a population prospectively would require criteria measured before allocation.

Finally, the patients were predominantly younger adults with severe injury, transported directly to French specialist centres. Treatment was early after hospital arrival, although substantial time had already elapsed after injury. The evidence does not directly cover children, transferred patients, traumatic arrest, pregnancy or patients taking anticoagulants. Within the studied setting, however, the practical implication is clear: rapid availability and favourable effects on a clotting assay are insufficient reasons to add PCC routinely to plasma-based resuscitation.

Body of Evidence

The preceding evidence addresses several distinct decisions: the balance of conventional components, replacement of plasma by a guided concentrate strategy, and addition of PCC to plasma. Keeping these questions separate explains why PROCOAG can challenge routine PCC use without overturning all factor-based approaches to traumatic bleeding. It also prevents an association observed with one resuscitation package from being attributed to a single ingredient.

PROPPR randomised 680 severely injured patients to plasma, platelets and red cells in ratios of 1:1:1 or 1:1:2. Mortality did not differ significantly at either 24 hours or 30 days, although haemostasis was achieved more often with 1:1:1 and death from

exsanguination was less frequent. This provides context for balanced component resuscitation and for the difficulty of choosing an endpoint in rapidly evolving haemorrhage. More early plasma or platelets may represent effective resuscitation rather than treatment failure. Conversely, fewer products cannot automatically be interpreted as better care. PROPPR did not evaluate PCC, but it shows why the clinical consequences of changing a transfusion strategy matter more than the product count alone.⁴

RETIC examined an alternative package in patients with bleeding and viscoelastic evidence of coagulopathy. In this single-centre, open-label trial, first-line coagulation-factor concentrates, primarily fibrinogen, were compared with plasma. The trial stopped early because rescue treatment was much more frequent with plasma. In the analysed population, multiple-organ failure occurred in 29/44 plasma recipients and 25/50 concentrate recipients, a non-significant primary comparison. Crucially, only eight of the 50 patients allocated to the concentrate strategy received PCC. RETIC therefore cannot establish that PCC itself was beneficial, nor does it answer whether PCC should be added empirically to plasma. Its selection process, treatment package, open-label design and early termination make a direct numerical comparison with PROCOAG inappropriate.⁵

A propensity-matched study by Jehan and colleagues compared 40 patients receiving 4F-PCC plus plasma with 80 receiving plasma alone. Participants had an international normalised ratio above 1.5 and were not taking anticoagulants before injury. PCC use was associated with lower red-cell and plasma requirements and more rapid correction of the international normalised ratio. Those findings supported further evaluation, but matching could address only measured differences. Treatment selection and contemporaneous resuscitation decisions remained possible explanations. The study's small number of thromboembolic events also could not establish safety with useful precision.⁶

A larger nationwide propensity-matched analysis by Zeeshan and colleagues used US trauma data from 2015–2016. It compared 234 patients in each group and associated PCC plus plasma with lower transfusion requirements and mortality than plasma alone. Reported mortality was 17.5% versus 27.7%. This was a substantial association, but its magnitude did not make it immune to confounding. Administrative and clinical registry data cannot fully reconstruct why a treatment was chosen, how quickly bleeding control occurred or all aspects of resuscitation. PROCOAG's concealed allocation consequently provides stronger evidence for the causal effect of routine supplementation than either observational comparison, even though the patient populations and systems were not identical.⁷

ITACTIC tested whether augmenting major-haemorrhage protocols with viscoelastic rather than conventional coagulation testing improved outcomes. Among 396 analysed patients, 67% versus 64% were alive and free of massive transfusion at 24 hours, odds ratio 1.15, 95% confidence interval 0.76–1.73. Twenty-eight-day mortality was also similar. This did not demonstrate an overall benefit from the alternative monitoring strategy, but it did not isolate the effect of each treatment triggered by that strategy. It neither validates empirical PCC nor proves that a specific factor-guided intervention cannot work. PROCOAG likewise did not randomise patients according to a demonstrated PCC-responsive deficit.⁸

By the end of 2023, the most defensible synthesis was therefore specific. Observational enthusiasm for adding PCC to plasma had not survived a direct blinded randomised test, and a clinically concerning thromboembolic signal had emerged. Evidence supporting balanced components or selected concentrate algorithms addressed related but different questions. Future evaluation of PCC would need to define the intended biological deficit, establish the role of background plasma and measure both patient benefit and thrombotic harm. These unresolved possibilities should not dilute the immediate result for the routine empirical strategy that PROCOAG actually tested.

Should 4F-PCC be added routinely to plasma-based resuscitation for severe traumatic bleeding?

PROCOAG does not support routine addition: it did not reduce transfusion requirements and increased detected thromboembolic events. Targeted use for a demonstrated deficit or anticoagulant reversal was not tested.

References

1. Bouzat P, Charbit J, Abback PS, Huet-Garrigue D, Delhay N, Leone M, et al.; PROCOAG Study Group. Efficacy and Safety of Early Administration of 4-Factor Prothrombin Complex Concentrate in Patients With Trauma at Risk of Massive Transfusion: The PROCOAG Randomized Clinical Trial. *JAMA*. 2023;329(16):1367–1375.
2. PROCOAG investigators. Statistical analysis plan, 8 April 2022. Supplement 2 to the PROCOAG randomised clinical trial. *JAMA*. 2023.
3. PROCOAG Study Group. Supplement 3 to: Efficacy and Safety of Early Administration of 4-Factor Prothrombin Complex Concentrate in Patients With Trauma at Risk of Massive Transfusion. *JAMA*. 2023;329:1367–1375.

4. Holcomb JB, Tilley BC, Baraniuk S, Fox EE, Wade CE, Podbielski JM, et al.; PROPPR Study Group. Transfusion of plasma, platelets, and red blood cells in a 1:1:1 vs a 1:1:2 ratio and mortality in patients with severe trauma: the PROPPR randomized clinical trial. *JAMA*. 2015;313(5):471–482.
5. Innerhofer P, Fries D, Mittermayr M, Innerhofer N, von Langen D, Hell T, et al. Reversal of trauma-induced coagulopathy using first-line coagulation factor concentrates or fresh frozen plasma (RETIC): a single-centre, parallel-group, open-label, randomised trial. *Lancet Haematol*. 2017;4(6):e258–e271.
6. Jehan F, Aziz H, O’Keeffe T, Khan M, Zakaria ER, Hamidi M, et al. The role of four-factor prothrombin complex concentrate in coagulopathy of trauma: A propensity matched analysis. *J Trauma Acute Care Surg*. 2018;85(1):18–24.
7. Zeeshan M, Hamidi M, Feinstein AJ, Gries L, Jehan F, Sakran J, et al. Four-factor prothrombin complex concentrate is associated with improved survival in trauma-related hemorrhage: A nationwide propensity-matched analysis. *J Trauma Acute Care Surg*. 2019;87(2):274–281.
8. Baksaas-Aasen K, Gall LS, Stensballe J, Juffermans NP, Curry N, Maegele M, et al. Viscoelastic haemostatic assay augmented protocols for major trauma haemorrhage (ITACTIC): a randomized, controlled trial. *Intensive Care Med*. 2021;47(1):49–59.

Possible Reporting Inconsistencies. Main Table 2 gives hospital-free-day medians and interquartile ranges identical to those for ICU-free days: 6.5 (0–22.5) with PCC and 7 (0–22) with placebo. Supplementary eTable 3 instead reports hospital-free days as 0 (0–0) in both groups in the large per-protocol subset. The SAP defines hospital-free and ICU-free days separately; the hospital-free-day summaries remain difficult to reconcile.^{1,2,3}

TOP-ART

Shepherd JM, Ross J, Anton L, Rourke C, Brentnall AR, Tarning J, et al. Safety and efficacy of artesunate treatment in severely injured patients with traumatic hemorrhage. The TOP-ART randomized clinical trial. *Intensive Care Med.* 2023;49(8):922–933.

Introduction

Survival from traumatic haemorrhage does not end the threat posed by severe injury. Patients may require prolonged support for respiratory, cardiovascular, renal and other organ dysfunction despite control of bleeding. Resuscitation addresses perfusion and haemostasis, but it does not necessarily prevent the subsequent biological response that contributes to organ injury. A treatment capable of reducing that response without impairing recovery would address a different clinical need from another haemostatic agent.

Artesunate offered an unusual opportunity for drug repurposing. It was already an established treatment for severe malaria, with extensive human exposure and a formulation suitable for rapid intravenous administration. Experimental work suggested effects on inflammation and cell-survival pathways beyond its antiparasitic action. The TOP-ART investigators had observed less organ injury after artesunate in a rat model of haemorrhage and resuscitation. Human studies of the earliest response to injury also suggested that processes associated with subsequent organ dysfunction were active within hours, providing a rationale for intervention during initial trauma care.¹

These observations were encouraging but left several important uncertainties. Pharmacological activity in a controlled animal model does not establish that the same dose and timing will modify human organ dysfunction. Nor does tolerability in malaria guarantee safety in patients with tissue injury, haemorrhage and a changing coagulation state. The relevant exposure for an anti-inflammatory or organ-protective effect might differ from that required to treat parasitaemia. An early clinical study therefore needed to examine safety alongside physiological efficacy.

TOP-ART tested whether a single early intravenous dose of artesunate could reduce organ dysfunction after severe traumatic bleeding. It was a phase IIa study intended to identify a treatment signal and establish whether progression was justified, rather than a definitive mortality trial. Its design brought the translational hypothesis into a

recognisable clinical population, but also required careful interpretation of a short-term organ score, two sequential dose stages and potentially uncommon adverse effects.

Synopsis

TOP-ART was a randomised, placebo-controlled, blinded phase IIa trial at the Royal London Hospital major trauma centre. Recruitment took place from 21 March 2017 to 15 May 2019, and follow-up ended on 10 August 2019. The Wellcome Trust and Department of Health funded the research. Sigma Tau supplied good-manufacturing-practice artesunate without charge. Recruitment was interrupted for several months in 2017 while a new batch of diluent was awaited, and again in 2018 pending interim review.¹

Patients aged at least 16 years were eligible after activation of the local massive-haemorrhage pathway. Activation required systolic pressure below 90 mmHg, suspected haemorrhage and minimal response to small-volume fluid resuscitation. Patients had to arrive within two hours of injury and be able to receive treatment within four hours. Exclusions included severe traumatic brain injury, expected death before the 48-hour efficacy assessment, pregnancy or breastfeeding, non-haemorrhagic shock, pathway activation more than an hour after arrival, concurrent participation in another investigational-drug trial and known artesunate allergy. These criteria selected patients at risk of organ dysfunction who were nevertheless expected to survive its initial assessment window.

For patients lacking capacity, agreement was obtained from a relative when appropriate or from two doctors, one independent of the trial. Continuing consent was sought when patients recovered capacity, or from an appropriate representative if they did not. Randomisation used varying block sizes and a 2:1 artesunate:placebo ratio. The trial statistician generated the sequence, an external statistician validated it, and an independent provider prepared sealed allocation cards. The decision to enrol preceded opening the next card. Participants, clinical staff and investigators collecting outcome data were blinded. The investigator preparing and administering treatment was unblinded because reconstitution required visual inspection.

Treatment was a single intravenous bolus within four hours of injury. The first stage compared artesunate 2.4 mg/kg with placebo. Following interim safety review, the second stage used 4.8 mg/kg. Artesunate powder was reconstituted immediately before administration in phosphate-buffer diluent at 10 mg/mL; placebo consisted of an equivalent volume of diluent. The two doses were therefore introduced sequentially, rather than concurrently randomised against one another. The report describes this as an adjunct to trauma care but does not provide a sufficiently detailed common resuscitation

algorithm to specify identical transfusion, operative or intensive-care management across all participants.

The primary efficacy endpoint was the Sequential Organ Failure Assessment (SOFA) score 48 hours after admission, with a lower score favouring artesunate. Serious adverse events during the first 28 days formed the safety endpoint. SOFA was recorded daily through day seven, and multiple organ dysfunction syndrome (MODS) was defined as any daily score above five. Secondary outcomes included maximum SOFA, mean SOFA during days three to five, organ-failure resolution, ventilator-free days, lengths of stay, acute lung and kidney injury, infection, and mortality at 28 days, discharge and 90 days. Persistent MODS required a score above five on day seven. An additional comparison assessed being alive and free of MODS at 48 hours.

The planned primary analysis was per protocol. A two-point reduction in 48-hour SOFA, with a standard deviation of 3.43, required 70 artesunate and 35 placebo patients for 80% power at a significance level of 0.05. This was a target of 105 evaluable patients, not simply 105 randomisations. Efficacy was also examined in an intention-to-treat population. Safety analyses followed treatment received. The investigators used Mann–Whitney and chi-squared comparisons, with empirical bootstrap confidence intervals for mean differences. Low-dose and high-dose comparisons were also presented. No formal statistical stopping boundary was specified in the main report; an independent monitoring committee reviewed safety repeatedly and advised the steering committee about continuation.

Of 211 massive-haemorrhage activations, 105 met eligibility criteria and 90 patients were randomised: 59 to artesunate and 31 to placebo. Fifteen eligible patients were missed, commonly because recruitment staff were unavailable. Five patients subsequently declined consent, leaving 85 in the stated intention-to-treat population. The principal safety tables included 54 artesunate recipients and 29 controls. The per-protocol analysis included 48 and 27, respectively, comprising 31 low-dose and 17 high-dose recipients. Exclusions included ineligibility, treatment deviations, death before 48 hours and early discharge. One patient allocated placebo received low-dose artesunate in error. The safety denominators used here follow the main and supplementary safety tables.²

The per-protocol population was young, with median age 29 years, median Injury Severity Score 22 and penetrating injury in 51%. Eighty-eight per cent were men, but all 27 controls were male compared with 39/48 artesunate recipients. Admission base deficit was greater with artesunate, median 9 versus 4.7 mmol/L, and Glasgow Coma Scale score was lower, median 14 versus 15. Median injury-to-admission times were 71

versus 88 minutes. Pharmacokinetic observations in seven patients were described as within ranges previously observed during malaria treatment; the main report does not supply a numerical distribution of actual injury-to-dose times.

Recruitment stopped after concern about thromboembolic events emerged during serial safety reviews. Four such serious events had occurred in one treatment group at the interim assessment. Further review, including an additional analysis after 90 randomisations, led the steering committee to recommend termination because of the possible safety imbalance and lack of an efficacy signal. This left 75 per-protocol participants, below the intended 105.

Median 48-hour SOFA was 5.5 with artesunate and four with placebo; the corresponding means were 6.4 and 5.3. The published mean difference, expressed as placebo minus artesunate, was -1.1 points, 95% confidence interval -3.1 to 0.9 ; $P=0.303$ for the rank comparison. Neither dose-specific comparison showed benefit. MODS developed in 33/48 versus 18/27 patients, and 24/48 versus 18/27 were alive and MODS-free at 48 hours. Median critical care stay was eight versus 11 days, $P=0.736$. Per-protocol mortality was two versus one at each reported follow-up point; these figures exclude patients removed before the primary analysis and are not mortality rates for all randomised patients.

Serious adverse events affected 17/54 artesunate recipients and 5/29 controls, relative risk 1.8, 95% confidence interval 0.8–4.4. There were 23 versus eight events. The subsequent detailed thromboembolic analysis identified nine versus one affected patients, relative risk 4.8, 95% confidence interval 0.6–36.3; $P=0.078$. The category termed venous thromboembolism included some other vascular events. Early low-molecular-weight heparin prophylaxis was documented in 18/51 versus 15/28 patients. Post hoc adjustment did not identify a treatment effect on SOFA. Supplementary Table 7a gives an available-case intention-to-treat comparison: 48-hour SOFA was observed in 50/54 artesunate patients and 29/31 controls, with medians of five and four and means of 6.3 and 5.1. The placebo-minus-artesunate mean difference was -1.2 points, 95% confidence interval -3.1 to 0.7 ; $P=0.211$. Table 7b records four versus two deaths in the intention-to-treat population at 28 days, discharge and 90 days. These counts differ from the treatment-received safety population and should retain their own denominators.²

Critique

TOP-ART did not provide a reason to introduce artesunate for prevention of organ dysfunction after traumatic haemorrhage. The primary efficacy comparison favoured

neither pooled treatment nor either individual dose, while a clinically concerning imbalance in thromboembolic events prompted termination. This is an appropriate basis for declining routine use. It is a less secure basis for claiming that artesunate definitively causes thrombosis or that all potentially useful effects have been excluded. Both efficacy and safety conclusions depend on a small, selected and incompletely recruited population.

The study had several practical strengths. Allocation was concealed until the enrolment decision, outcome collectors and treating clinicians were blinded, and placebo matched the administration procedure. These protections matter when outcomes incorporate treatment-dependent physiological measurements and decisions about organ support. Recruiting during emergency trauma care and obtaining early pharmacokinetic samples also demonstrated that the proposed translational programme was feasible. The unblinded drug administrator represents a potential route for disclosure, but the report does not establish that blinding failed.¹

The primary analysis population is a more substantive concern. Patients dying before 48 hours were excluded from the per-protocol efficacy analysis, despite death being an essential outcome for an intervention intended to protect organs. An early discharge also led to exclusion. The resulting score comparison therefore concerns a selected group surviving and remaining assessable at the specified time. Removing patients at both extremes of recovery can change the apparent distribution of organ dysfunction. A low score among survivors cannot by itself demonstrate that a treatment improves the overall trajectory of all patients offered it.

The supplementary intention-to-treat results partly address this problem, but the SOFA comparison remains an available-case analysis. The SAP prespecified separate summaries of missing 48-hour scores and of scores among those with measurements. Thus the intention-to-treat label does not mean that deaths or other missing scores were assigned a value in that comparison: six of the 85 patients lacked a 48-hour score. This is an explicit analysis choice rather than a reporting error, but it still does not estimate an outcome incorporating death for every randomised patient. The supplementary mortality counts and alive-and-MODS-free comparison, 28/54 versus 21/31, should be considered alongside SOFA. The low per-protocol mortality remains especially selective because of both the initial exclusion of patients expected to die and subsequent removal of early deaths.^{2,3}

SOFA is useful for quantifying physiological dysfunction but is not a direct measure of how patients feel, function or survive. Different organ components can contribute equally to its total while having different clinical consequences. A single 48-hour

observation may also miss a treatment that alters later persistence rather than initial severity. The trial included trajectories and persistent MODS outcomes, which help address that limitation, and these did not reveal a convincing alternative benefit. Nevertheless, the absence of a score improvement cannot establish that every relevant inflammatory pathway was unaffected.

Early termination reduced precision, but the direction and interval of the primary estimate deserve attention. In the per-protocol comparison, artesunate had a higher mean SOFA. The published interval for placebo minus artesunate did not extend to the planned two-point benefit. This is more informative than merely noting $P > 0.05$. It still does not create an equivalence trial or remove the selection issues in the analysis population. The rank-test probability and bootstrap mean-difference interval also describe different statistical summaries; the 1.5-point difference between displayed medians should not replace the reported mean effect.

Chance imbalance was substantial. Artesunate recipients had more marked metabolic evidence of shock at admission, an important prognostic difference in a trial of post-haemorrhagic organ dysfunction. All controls in the primary analysis were male. Small randomised groups need not be balanced in every characteristic, and these differences do not prove defective randomisation. They do, however, make the realised comparison less stable. The post hoc regression adjusting for age, sex, injury severity and base deficit did not uncover efficacy, but a model fitted to this small dataset cannot provide the assurance of a larger prospectively adjusted analysis.

The safety findings require similar care. Nine versus one patients is a striking imbalance, yet the relative-risk interval ranges from 0.6 to 36.3. This includes no increase and very substantial harm. Statistical non-significance is not evidence of safety, particularly when the intervention has no established benefit. Equally, the point estimate should not be described as a precise fivefold increase. A monitoring committee can reasonably stop an exploratory trial over a concerning pattern without having proved causation at a conventional statistical threshold.

The thromboembolic analysis also developed after the accumulating adverse-event pattern had been seen. It was not an independently powered primary safety comparison. No routine screening programme was used, and events were detected clinically. The category labelled venous thromboembolism included a cerebral arterial infarction and an embolism identified during injury repair, as well as venous thrombosis and pulmonary embolism. This broad grouping is relevant to overall vascular safety but should not be translated into nine patients with conventional venous thromboembolism. Event counts and numbers of affected patients must also remain separate.

Fewer artesunate recipients received heparin prophylaxis within 48 hours, and their greater shock could plausibly influence both thrombosis and the decision to delay prophylaxis. These are competing explanations, not established corrections to the result. Prophylaxis is a postrandomisation exposure and may itself be affected by illness or treatment. Adjusting for it can introduce additional bias or remove part of a treatment pathway. With only ten affected patients overall, multivariable safety modelling is particularly fragile. It cannot demonstrate that prophylaxis explains the signal or that artesunate is harmless after adjustment.

Sequential dose stages further limit interpretation. The larger number of events in the low-dose group does not establish an inverse dose response or show that the higher dose is safe. Only 20 high-dose recipients entered the safety analysis, the stages occurred at different times, and pooled placebo comparisons do not recreate a direct randomised dose comparison. Likewise, achieving concentrations familiar from malaria supports drug delivery but does not verify target engagement for organ protection after trauma.

Finally, this was one specialist centre with many young men and penetrating injuries. Severe brain injury and patients expected to die early were excluded. The result has limited direct applicability to older, predominantly blunt-trauma populations or the full spectrum of haemorrhagic shock. Within its proper scope, however, the study is useful: it tested a plausible repurposing hypothesis, found no convincing clinical efficacy signal and identified a safety concern that justifiably restrained further routine use.

Body of Evidence

The evidence leading to TOP-ART came from three different sources: established treatment of malaria, experimental haemorrhagic shock and observational characterisation of human organ dysfunction. Each supported a part of the rationale, but none had already established that artesunate prevented organ failure in injured patients. The trial's importance lies in testing that final clinical step rather than in disproving all of the preceding biological observations.

In SEAQUAMAT, 1,461 patients with severe falciparum malaria were randomised to artesunate or quinine. Mortality was 107/730 with artesunate and 164/731 with quinine. Artesunate was well tolerated, while hypoglycaemia was more frequent with quinine. These results established a substantial clinical benefit in severe malaria and made repurposing an existing drug more attractive than beginning with an entirely unfamiliar compound. However, the treatment regimen involved repeated dosing and the principal therapeutic target was infection. It did not demonstrate an organ-protective effect

independent of parasite clearance, nor establish vascular safety in traumatic haemorrhage. Malaria experience justified investigation, but could not substitute for trauma-specific randomisation.⁴

Sordi and colleagues supplied the more direct experimental bridge. Rats underwent controlled haemorrhagic shock, with mean arterial pressure reduced to 30 mmHg for 90 minutes, followed by resuscitation and random allocation to artesunate or vehicle. Doses of 2.4 or 4.8 mg/kg attenuated organ injury assessed four hours later. Molecular analyses implicated activation of cell-survival signalling and inhibition of inflammatory pathways. These findings provided a coherent mechanism and informed the clinical doses. Their limitations for translation are nevertheless substantial: controlled shock, a short outcome window and experimental organ-injury measures differ from heterogeneous injuries and complex human resuscitation. TOP-ART did not reproduce the expected improvement in its clinical score.⁵

Cabrera and colleagues examined the human response within two hours of injury using transcriptomic and flow-cytometry approaches. Their findings associated later MODS with an early pattern involving cell-survival and cell-death pathways. The 2018 correction clarified that the analysed transcriptomic comparison included 28 critically injured patients, of whom 12 developed MODS, rather than the larger numbers initially described. This supports interest in the earliest postinjury period but also emphasises the small observational foundation. An association between gene expression and later dysfunction does not identify which pathway should be modified, whether it is harmful or adaptive, or whether artesunate changes it sufficiently in humans.^{6,7}

Subsequent observational work also showed that post-trauma organ dysfunction has different recovery patterns. The ORDIT study described early resolving and more persistent forms, with substantial mortality and resource use in patients developing MODS. This is relevant to interpretation of a treatment assessed primarily at 48 hours: the same early aggregate score may precede different subsequent courses. It does not establish a responsive subgroup for artesunate, but it supports attention to the duration and pattern of dysfunction as well as its initial severity. The clinical classifications remain descriptions of outcome trajectories, rather than proof of discrete drug-responsive mechanisms.⁸

By the 2023 evidence cut-off, TOP-ART had therefore weakened the case for the specific early single-dose strategy without resolving the wider biology of organ dysfunction. The preclinical work and human observational studies remained hypothesis-generating; the successful malaria trials remained applicable to malaria. The absence of an efficacy signal, combined with uncertain but potentially important vascular harm, made routine

trauma use unjustified. Further investigation would require a more persuasive account of human target engagement, an explicit strategy for incorporating death into efficacy assessment and a safety design capable of distinguishing treatment-related thrombosis from the background consequences of injury and its management.

Should early intravenous artesunate be used to prevent organ dysfunction after traumatic haemorrhage?

TOP-ART does not support this use: organ dysfunction did not improve, and a possible increase in thromboembolic events prompted early termination. The small trial cannot establish the precise magnitude or cause of that safety signal.

References

1. Shepherd JM, Ross J, Anton L, Rourke C, Brentnall AR, Tarning J, et al. Safety and efficacy of artesunate treatment in severely injured patients with traumatic hemorrhage. The TOP-ART randomized clinical trial. *Intensive Care Med.* 2023;49(8):922–933.
2. Shepherd JM, Ross J, Anton L, et al. Supplementary information to: Safety and efficacy of artesunate treatment in severely injured patients with traumatic hemorrhage. The TOP-ART randomized clinical trial. *Intensive Care Med.* 2023;49:922–933.
3. TOP-ART investigators. Statistical analysis plan, version 2, 6 August 2018. Supplementary material to the TOP-ART randomised clinical trial.
4. Dondorp A, Nosten F, Stepniewska K, Day N, White N; South East Asian Quinine Artesunate Malaria Trial (SEAQUAMAT) group. Artesunate versus quinine for treatment of severe falciparum malaria: a randomised trial. *Lancet.* 2005;366(9487):717–725.
5. Sordi R, Nandra KK, Chiazza F, Johnson FL, Cabrera CP, Torrance HD, et al. Artesunate Protects Against the Organ Injury and Dysfunction Induced by Severe Hemorrhage and Resuscitation. *Ann Surg.* 2017;265(2):408–417.
6. Cabrera CP, Manson J, Shepherd JM, Torrance HD, Watson D, Longhi MP, et al. Signatures of inflammation and impending multiple organ dysfunction in the hyperacute phase of trauma: A prospective cohort study. *PLoS Med.* 2017;14(7):e1002352.

7. Cabrera CP, Manson J, Shepherd JM, Torrance HD, Watson D, Longhi MP, et al. Correction: Signatures of inflammation and impending multiple organ dysfunction in the hyperacute phase of trauma: A prospective cohort study. PLoS Med. 2018;15(10):e1002694.
8. Cole E, Gillespie S, Vulliamy P, Brohi K; Organ Dysfunction in Trauma (ORDIT) study collaborators. Multiple organ dysfunction after trauma. Br J Surg. 2020;107(4):402–412.

Possible Reporting Inconsistencies. Figure 1 appears to contain treatment-receipt and safety-population errors: it records all 59 artesunate allocations as receiving treatment despite footnotes describing non-recipients, and gives 30 placebo patients in the safety population. The main safety tables and supplementary Table 2 consistently use 29 placebo patients. This chapter uses those tabulated safety denominators; differences between intention-to-treat and safety mortality counts concern different populations.^{1,2}



Paediatric Trials

OXY-PICU

Peters MJ, Gould DW, Ray S, Thomas K, Chang I, Orzol M, et al.; Oxy-PICU Investigators of the Paediatric Critical Care Society Study Group (PCCS-SG). Conservative versus liberal oxygenation targets in critically ill children (Oxy-PICU): a UK multicentre, open, parallel-group, randomised clinical trial. Lancet. 2024;403(10424):355–364.

Introduction

Supplemental oxygen is central to the management of respiratory failure in children, but its optimum dose is uncertain. A high peripheral oxygen saturation offers reassurance that hypoxaemia has been corrected, while giving little indication of how much oxygen might safely be withdrawn. Once haemoglobin is nearly fully saturated, additional oxygen can substantially increase arterial oxygen tension without a useful change on the bedside saturation display. Avoiding hypoxaemia and avoiding unnecessary oxygen exposure are therefore related but distinct treatment goals.¹

The rationale for conservative oxygenation includes reducing oxidative injury and limiting treatment delivered simply to sustain a high saturation. The countervailing concern is that a lower target reduces the margin for deterioration, particularly when oxygen delivery is compromised by circulatory failure or anaemia. Neither physiological argument identifies a universally safe threshold. Age, underlying disease, the reliability of monitoring and the clinical response to a falling saturation may all influence the consequences of adopting a lower target.

Paediatric evidence had supported further investigation without resolving the question for ventilated children. The pilot Oxy-PICU trial demonstrated that clinicians could achieve different oxygen exposures, while the BIDS trial examined a lower oxygen threshold in infants hospitalised with bronchiolitis. These populations and objectives differed from a definitive intensive care effectiveness trial. Evidence from adults and extremely premature infants also required careful separation from the question facing a general paediatric intensive care unit (PICU).^{2,3}

Oxy-PICU tested whether a conservative oxygenation strategy could reduce organ support or death among children admitted urgently to PICU and receiving invasive ventilation. Its practical question concerned assigning a bedside target with clinical discretion, rather than imposing a fixed arterial oxygen tension or eliminating every excursion outside a saturation range. Although its journal issue appeared in January

2024, the principal report was published online on 1 December 2023 and falls within this book's evidence period.¹

Synopsis

Oxy-PICU was a pragmatic, open, parallel-group superiority trial conducted in 15 UK National Health Service PICUs, with recruitment also undertaken by participating transport services. Enrolment ran from September 2020 to May 2022. The National Institute for Health and Care Research funded the study, sponsored by the Intensive Care National Audit & Research Centre. The funder had no stated role in the design, analysis or publication decision.¹

Eligible children were older than 38 weeks corrected gestational age and younger than 16 years, admitted as an emergency and receiving invasive ventilation with supplemental oxygen for abnormal gas exchange. Randomisation was required within six hours of meeting the criteria. Important exclusions included a primary brain pathology or injury, pulmonary hypertension, sickle cell disease, uncorrected congenital cardiac disease, established long-term invasive ventilation, treatment limitation and predicted death before reaching intensive care. Recruitment during a previous admission also precluded enrolment.

Allocation was concealed through a central telephone or web system. Children were assigned equally using minimisation with an 80% probability of allocation to the arm that best balanced age, centre, lower respiratory tract infection and the combination of saturation-to-inspired-oxygen ratio with positive end-expiratory pressure. Treatment was unmasked. Research could begin without prior consent; parents were approached subsequently when appropriate, usually within 24–48 hours. Refusal or withdrawal could therefore remove a child from the analysed population after randomisation.^{1,4}

The conservative group targeted peripheral oxygen saturation of 88–92%, compared with greater than 94% in the liberal group. The assigned strategy started immediately and continued during invasive ventilation, including restarting it after failed extubation or readmission before day 30. Oxygen concentration and ventilator settings could be adjusted, with clinical decisions otherwise left to treating teams. A saturation above the conservative range while breathing room air did not require an intervention to lower it. Clinicians could override allocation when necessary; this was a comparison of implementable management strategies rather than mandatory continuous exposure to the nominal target.

The primary endpoint ranked the calendar days with any organ support through day 30, beginning on the randomisation day. Survivors received their observed support-day value; death by day 30 received the worst rank, 31. Lower values were favourable. Support included invasive and non-invasive ventilation, cardiovascular and renal therapies, but also interventions such as sedative infusions, blood transfusion and neurological support. Supplemental oxygen alone was not respiratory organ support. Relevant support outside the initial PICU stay was included, so the endpoint was broader than ventilation duration or PICU length of stay.⁴

Secondary outcomes included mortality, sustained liberation from invasive ventilation, individual support measures, lengths of stay, discharge functional categories and 30-day costs. Liberation required 48 continuous hours without invasive ventilation. The planned sample of 2,040 allowed 90% power to detect a 12-hour reduction from an assumed mean 120 hours of organ support, with assumed mortality of 7.5%, a two-sided 5% significance level and 10% loss associated with consent. An interim review at approximately half recruitment used a stringent Peto–Haybittle boundary; the trial completed recruitment.

The primary comparison used the Wilcoxon rank-sum test and a probabilistic index, with 0.5 representing no distributional advantage. Ordered logistic regression provided unadjusted and adjusted odds ratios, below one favouring conservative oxygenation. Adjustment incorporated age, respiratory diagnosis, oxygenation severity, predicted mortality and centre. Missing support status was inferred from care location for 60 children: support was assumed during intensive or high-dependency care, or when location was unknown, and absent on a ward. Secondary analyses were not protected by a multiplicity adjustment.^{1,4}

Of 6,089 children screened, 2,817 were eligible and 2,040 were randomised: 1,022 to conservative and 1,018 to liberal oxygenation. After consent-related exclusions and one duplicate allocation, 939 and 933 remained in the primary analysis. The median age was approximately one year, almost half were younger than 12 months, and lower respiratory tract infection accounted for 64% and 65% of admissions. Around half had a recorded comorbidity. Mean predicted mortality was 3.7% and 3.8%, and approximately 38% were randomised through transport services. Baseline median saturation was 98% in both groups.

Treatment assignment produced clear separation, although achieved saturations were above the conservative target much of the time. The median of individual mean saturations was 94% versus 97%; corresponding inspired oxygen fractions were 0.27 versus 0.35. Median time within 88–92% was 26% versus 1%, while conservative children

commonly remained above target on room air. During the first seven days, deviations affected 389/939 versus 155/933 children but occupied only 3.8% versus 1.5% of recorded hours. Mean proportions of time below 88% were 1.47% and 0.54%.⁴

The primary outcome favoured conservative oxygenation: median support duration was five days in both groups, with interquartile ranges of 3–9 and 3–10 days. The probabilistic index was 0.53 (95% CI 0.50–0.55; $P=0.04$), the unadjusted odds ratio 0.85 (95% CI 0.73–0.99) and adjusted odds ratio 0.84 (95% CI 0.72–0.99). By day 30, 23/939 children versus 28/933 had died; the adjusted odds ratio was 0.80 (95% CI 0.45–1.41; $P=0.48$). There was no demonstrated mortality reduction.¹

Median time to liberation from invasive ventilation was 69.6 versus 73.2 hours, with an adjusted hazard ratio of 1.11 (95% CI 1.01–1.21; $P=0.03$). PICU and hospital stays did not differ convincingly. Mean 30-day costs were £32,479 versus £34,725; the adjusted difference of –£2,143 had a confidence interval from –£4,320 to £34. Discharge functional categories were similar, but baseline functional scores were not collected. Serious adverse events affected 12/939 versus 17/933 children. Per-protocol results were similar to the primary analysis; restricting support to respiratory, cardiovascular and renal components yielded an adjusted odds ratio of 0.86 (95% CI 0.73–1.01). No convincing interaction emerged across the reported age, haemoglobin or heart-rate subgroups.^{1,4}

Critique

Oxy-PICU provides evidence for a modest improvement in a combined support-duration and mortality ranking when eligible children are assigned a conservative oxygen strategy. It does not demonstrate that maintaining every child continuously at 88–92% improves survival. That distinction is central: the median mean saturation in the conservative group was 94%, and substantial time above target reflected room-air breathing. The clinically tested intervention was willingness to reduce oxygen and tolerate a lower saturation, embedded in bedside decision-making. A policy that actively forces saturations into the lower range could produce a different exposure and safety profile.

Several design features support confidence in the comparison. Central concealment protected allocation, individual randomisation avoided the interpretative difficulties of a small number of clusters, and minimisation balanced important prognostic factors. Recruitment through transport services enabled treatment early in the illness pathway. The multicentre sample included children with comorbidities and reflected a common

PICU presentation. Treatment separation was demonstrated directly, so the modest result cannot simply be dismissed as a trial in which both arms received identical oxygen.

Nevertheless, nominal adherence and physiological exposure are not interchangeable. Although more than 40% of conservative children had at least one recorded deviation, deviations occupied a small fraction of observed time. Conversely, remaining outside 88–92% on room air was compatible with the strategy. Removing every child with an excursion would create a selected comparison. The reported per-protocol analysis excluded specified eligibility or immediate treatment departures, rather than establishing the effect of uninterrupted target attainment. Clinical overrides might also have prevented harm. The trial does not show that tighter enforcement would enlarge the benefit.⁴

The primary endpoint sensibly ranked death below all survival outcomes, avoiding an apparent advantage from a short fatal admission. However, a rank of 31 is a statistical convention, not 31 observed days of organ support. The result combines the probability and duration of multiple therapies with mortality and cannot be translated into a fixed number of support hours saved for each child. Identical medians conceal small shifts across the distributions. A probabilistic index of 0.53 describes a slight distributional advantage; it is neither a 53% individual chance of benefiting nor the probability that the trial hypothesis is true.

The breadth of organ support also matters. Sedative infusions and other treatments may cease when a clinician becomes comfortable with respiratory progress, without necessarily reflecting prevention of intrinsic organ injury. Restricting the endpoint to respiratory, cardiovascular and renal support retained a similar direction but crossed the conventional significance boundary. This does not establish inconsistency or erase the primary result, but it qualifies the claim that all sensitivity analyses independently confirmed it. The more defensible interpretation is directional consistency with some dependence on endpoint definition and limited statistical precision.⁴

An unmasked oxygen trial is difficult to avoid, but here knowledge of the target can influence the outcome itself. Decisions about extubation, non-invasive support, sedation and discharge respond to clinicians' interpretation of an acceptable saturation. Such changes may represent a useful reduction in treatment burden, which a pragmatic trial should capture. They do not, however, prove that less oxygen prevented biological injury. Objective mortality was uncommon and imprecisely estimated, while most of the primary information came from support duration among survivors. The survival confidence interval remains compatible with clinically relevant benefit or harm.

Post-randomisation exclusions are another important qualification. Approximately 8% of allocated children did not enter the primary analysis, predominantly because of consent arrangements. Losses were numerically balanced, but balance does not establish that outcomes were similarly distributed. Supplementary best- and worst-case analyses assigned extreme outcomes to 114 children withdrawn or refused before day 30. These assumptions could reverse the primary result, producing a probabilistic index of 0.47 in the unfavourable scenario. The exercise does not show that such a scenario is plausible, and it does not cover every excluded child. It does show that the modest observed advantage is not insensitive to missing outcomes.⁴

The significant ventilation-liberation comparison is supportive but secondary. Its hazard ratio is not an 11% reduction in ventilation time. Deaths were censored in that analysis, unlike their worst-rank treatment in the primary endpoint, so it cannot by itself be read as an improvement in survival free of ventilation. Multiple secondary comparisons further reduce the weight of an isolated P value. Similarly, the cost confidence interval includes no saving, despite a rounded P value of 0.05. The economic result suggests potential savings over 30 days; it does not establish long-term cost effectiveness.

Safety requires a separate judgement. Serious adverse events were numerically fewer with conservative oxygenation, but small event counts cannot establish equivalence. Conservative children spent more time below 88%, even though the median number of such hours was zero. Similar discharge functional categories are reassuring only within their limits: baseline scores were unavailable, and this report did not establish long-term neurodevelopmental safety.^{1,4}

Applicability is strongest to the trial's predominantly young, respiratory-disease population. Children with primary brain injury, pulmonary hypertension, sickle cell disease and important uncorrected cardiac abnormalities were excluded. Extremely premature infants were outside the age range. These exclusions remove several groups in whom a lower oxygen target might have different consequences. Recruitment during the pandemic also occurred within an unusual pattern of respiratory admissions. Oxy-PICU supports considering a conservative strategy in comparable children with appropriate monitoring and clinical discretion. It does not define one optimum saturation for every PICU diagnosis, prove superiority over an intermediate target, or justify ignoring deterioration because a displayed value remains within 88–92%.

Body of Evidence

The 2018 pilot Oxy-PICU trial established the feasibility of the later programme. It randomised 119 children across three PICUs, with consent obtained for analysis in 107,

and included both invasive and non-invasive ventilation. Median time-weighted saturations were 94.9% and 97.5%, accompanied by lower inspired oxygen in the conservative arm. Clinical outcomes did not show an efficacy difference, but the study was designed to prepare a definitive comparison. Its contribution was to show that target assignment could change actual treatment, not to prove that a small trial without a clear harm signal had settled safety.²

BIDS addressed another paediatric decision: whether infants admitted with bronchiolitis needed oxygen to maintain a saturation of at least 94%. Among 615 infants aged six weeks to 12 months, modified oximeters allowed a 90% threshold to be compared with 94% while maintaining masking. Cough resolution was equivalent within the specified two-day margin, with a median of 15 days in both groups. This was evidence for a lower threshold in a defined ward population, not a test of invasive ventilation or an 88–92% PICU target. Its masking also reduced the direct visibility of allocation to clinicians.³

Taken together, these earlier studies supported investigating oxygen reduction while leaving the intensive care question open. A ward equivalence trial asks whether less oxygen preserves an acceptable symptomatic outcome. A PICU superiority trial asks whether assignment to a different target improves a combined treatment-burden and survival outcome. Their conclusions cannot be exchanged simply because both involve children and pulse oximeters. Oxy-PICU's added contribution is its larger, directly relevant randomised comparison during invasive ventilation and its demonstration of achieved exposure under routine clinical conditions.

Adult evidence reinforces the need to distinguish physiological separation from clinical benefit. In the 2022 PILOT trial, 2,541 mechanically ventilated adults were treated during cluster-crossover periods targeting saturations of 90%, 94% or 98%. Median ventilator-free days were 20, 21 and 21, respectively, without a significant overall difference. The trial included an intermediate strategy that Oxy-PICU did not examine. Its population had substantially higher mortality and different disease patterns, while its single-centre cluster design differed from individual paediatric randomisation. The result neither disproves a small paediatric benefit nor identifies which target is best for children.⁵

The BOOST II trials provide a different caution. Among extremely premature infants, targets of 85–89% were compared with 91–95%. The 2013 report found higher mortality with the lower target in the subgroup managed using the revised oximeter algorithm, alongside differing retinopathy and necrotising-enterocolitis outcomes. The infants' gestational age, target ranges and developmental vulnerability make direct extrapolation to Oxy-PICU inappropriate. The relevance is methodological and clinical:

reducing oxygen exposure can alter competing outcomes, and a favourable respiratory or treatment-burden measure cannot automatically establish overall safety.⁶

By the end of 2023, the evidence therefore supported a more discriminating approach to oxygen than treating the highest achievable saturation as the universal goal. Oxy-PICU supplied the strongest direct evidence within its enrolled population, but its effect was small and concerned a strategy whose achieved saturation was generally higher than the nominal conservative range. The primary ranking, secondary ventilation result and actual oxygen reduction point in a coherent direction, while mortality precision, missing outcomes and absent long-term function constrain the strength of the conclusion. The practical advance is evidence that selected ventilated children can be managed with less supplemental oxygen and may need slightly less organ support. The unresolved questions concern the safest implementation, the value of intermediate targets, and outcomes in the high-risk populations that the trial deliberately excluded.

Should ventilated children routinely receive a conservative oxygen target?

For children matching Oxy-PICU eligibility, a conservative strategy may modestly reduce organ support. Survival benefit and the safety of tightly maintaining every child at 88–92% remain unproven.

References

1. Peters MJ, Gould DW, Ray S, Thomas K, Chang I, Orzol M, et al.; Oxy-PICU Investigators of the Paediatric Critical Care Society Study Group (PCCS-SG). Conservative versus liberal oxygenation targets in critically ill children (Oxy-PICU): a UK multicentre, open, parallel-group, randomised clinical trial. *Lancet*. 2024;403(10424):355–364.
2. Peters MJ, Jones GAL, Wiley D, Wulff J, Ramnarayan P, Ray S, et al.; Oxy-PICU Investigators for the Paediatric Intensive Care Society Study Group (PICS-SG). Conservative versus liberal oxygenation targets in critically ill children: the randomised multiple-centre pilot Oxy-PICU trial. *Intensive Care Med*. 2018;44(8):1240–1248.
3. Cunningham S, Rodriguez A, Adams T, Boyd KA, Butcher I, Enderby B, et al.; BIDS group. Oxygen saturation targets in infants with bronchiolitis (BIDS): a double-blind, randomised, equivalence trial. *Lancet*. 2015;386(9998):1041–1048.
4. Oxy-PICU Investigators. Supplementary appendix to: Conservative versus liberal oxygenation targets in critically ill children (Oxy-PICU). *Lancet*. 2023.

5. 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.
6. Stenson BJ, Tarnow-Mordi WO, Darlow BA, Simes J, Juszczak E, Askie LM, et al.; BOOST II United Kingdom, Australia, and New Zealand Collaborative Groups. Oxygen saturation and outcomes in preterm infants. *N Engl J Med*. 2013;368(22):2094–2104.

Possible Reporting Inconsistencies. The Results state that all sensitivity analyses confirmed a statistically significant benefit. However, supplementary eTable 8 gives $P=0.07$ for the analysis restricted to respiratory, cardiovascular and renal support. Its direction favours the conservative group, but it does not meet the conventional 0.05 threshold; the blanket significance statement overstates the consistency of the sensitivity results.^{1,4}

SafeBoosC-III

Hansen ML, Pellicer A, Hyttel-Sørensen S, Ergenekon E, Szczapa T, Hagmann C, et al. Cerebral Oximetry Monitoring in Extremely Preterm Infants. *N Engl J Med*. 2023;388(16):1501–1511.

Introduction

The transition from fetal to neonatal life exposes the extremely preterm brain to rapid changes in oxygen delivery. Respiratory instability, immature cardiovascular regulation, anaemia and fluctuations in carbon dioxide can disturb the balance between oxygen supply and demand. Conventional arterial saturation and blood-pressure measurements describe parts of this physiology but do not directly measure oxygenation within the brain. A monitor capable of identifying cerebral hypoxia could therefore provide information that prompts earlier or more appropriate treatment.¹

Near-infrared spectroscopy estimates regional tissue haemoglobin oxygen saturation through a sensor placed on the head. The signal predominantly reflects venous blood in the underlying cortex and subcortex. A low value can prompt reassessment of circulation, respiratory support or oxygen-carrying capacity. However, monitoring only benefits an infant if the information identifies a reversible problem and the resulting treatment improves an outcome that matters. More measurements and more interventions are not intrinsically beneficial, and a physiological threshold is not automatically a validated treatment target.

The phase II SafeBoosC trial provided an encouraging intermediate result. Combining visible cerebral oximetry with a treatment guideline reduced the measured burden of abnormal cerebral oxygenation, principally hypoxia. Its size was insufficient to establish an effect on death or major brain injury. The subsequent challenge was to determine whether the approach could deliver clinical benefit when used more widely, across centres with differing devices, experience and usual care.²

SafeBoosC-III tested that next step. It examined cerebral oximetry linked to suggested clinical responses during the first 72 hours after birth in infants born before 28 weeks' gestation. The intended benefit was survival without severe brain injury by 36 weeks' postmenstrual age. This was a trial of adding a monitoring-and-treatment strategy to neonatal intensive care, rather than a comparison between oxygen saturation targets or a trial of the diagnostic accuracy of one particular sensor.¹

Synopsis

SafeBoosC-III was an investigator-initiated, pragmatic, open-label, multinational phase III superiority trial. Seventy sites in 17 countries across Europe, Asia and North America recruited infants between 27 June 2019 and 16 December 2021. Funding included the Capital Region of Denmark, charitable foundations and local institutional and commercial support. Some centres received equipment or technical assistance from device suppliers. The published appendix states that local funding sources had no role in design, analysis, interpretation or writing.^{1,3}

Eligible infants were born before 28 weeks' gestation, had a decision for full life support and could start cerebral oximetry within six hours of birth. Gestational age was assessed according to local practice. Prior written parental consent was used most commonly, but locally approved deferred-consent and opt-out approaches were permitted. Recruitment logs were not required, so the full screened and eligible populations were unavailable. Data from 45 centres suggested that enrolment represented, on average, 56% of their extremely preterm admissions.

Central computer-generated randomisation used varying block sizes and equal allocation, stratified by site and gestational age below or at least 26 weeks. The sequence was concealed. Multiple-birth siblings were allocated to the same group; when monitor availability prevented inclusion of all siblings, the last-born infant or infants were selected. Parents and clinical teams knew allocation. Severe brain injury assessment was blinded through locally organised procedures, sometimes undertaken nationally, but other exploratory outcomes were unblinded. Statisticians, data managers and authors worked with masked group identities, and alternative abstracts were agreed before the code was revealed.^{1,3}

The intervention comprised continuous cerebral oximetry, initiated within six hours and continued until 72 hours after birth. A head sensor displayed regional oxygenation. Values below a device-specific hypoxic threshold prompted consideration of treatment using a shared guideline. The reference threshold was 55% with the INVOS small adult sensor; equivalent thresholds for other approved neonatal devices were derived using a blood-lipid phantom. These were not interchangeable raw numerical readings: the appendix lists thresholds ranging from 48% to 67% across sensor combinations.

The guideline encouraged assessment of cardiovascular status, oxygen transport and respiratory status. Depending on the findings, clinicians could consider volume expansion, adjustment of vasoactive treatment, management of a patent ductus arteriosus, red-cell transfusion, changes in inspired oxygen or airway pressure, and reduction of excessive ventilation when carbon dioxide was low. These were potential

responses, not a mandatory drug or transfusion regimen for every low reading. The strategy targeted hypoxia; it did not retain the phase II programme's upper cerebral oxygenation target. Usual-care infants received conventional monitoring and treatment without visible cerebral oximetry. Masked oximetry for other research was permitted.³

The primary endpoint was death or survival with severe brain injury by 36 weeks' postmenstrual age. Assessment used information available up to that age, earlier death or discharge home. Severe injury comprised grade 3 or 4 intraventricular haemorrhage, cystic periventricular leukomalacia, post-haemorrhagic ventricular dilatation, cerebellar haemorrhage or cerebral atrophy on routine ultrasonography. Available scans and imaging reports were reviewed; there was no central image adjudication or routine magnetic resonance endpoint. Information was sought after hospital transfer, with earlier site data used when subsequent records could not be obtained.¹

Exploratory outcomes included bronchopulmonary dysplasia, severe retinopathy, late-onset sepsis and necrotising enterocolitis or focal intestinal perforation. These were presented post hoc as composites with death to address competing mortality, with the original separate outcomes retained in the appendix. The primary and exploratory components constituted serious adverse events. Device-related serious reactions included severe skin injury, displacement of tubes or vascular lines, and harmful clinical management prompted by the monitoring.

A sample of 1,600 was planned for 90% power to reduce the primary event rate from 34.0% to 26.5%, a 22% relative reduction, at a two-sided 5% significance level. One planned interim assessment after approximately one third of the sample used sequential monitoring boundaries and recommended continuation. Primary analyses followed allocation in infants with available outcomes. Mixed-effect logistic models included site as a random effect and treatment and gestational-age category as fixed effects; results were presented as relative risks. Two analysts worked independently before unmasking. Sensitivity analyses considered adherence, site effects, sibling dependence and missing data; exploratory intervals were not adjusted for multiple testing.^{1,3}

Of 1,601 randomised infants, 785 were assigned to cerebral oximetry and 816 to usual care. Primary data were available in 772 and 807, respectively, or 98.6% overall. Eighteen infants lacked eventual parental consent, two had permission to use data withdrawn and two were lost when a site withdrew. Median gestational age was 26.1 weeks in both groups, and median birth weight was 815 versus 814 g. Approximately one quarter were twins or triplets, and just over half were male. About 79% and 78% received invasive ventilation during their clinical course.

Monitoring began at a median age of three hours. INVOS neonatal sensors were used in 55.7% and OxyPrem devices in 30.2% of monitored infants. Thirty-six of 772 had more than 14 hours of missing monitoring, while 20/807 controls received visible cerebral oximetry. Including masked recordings, 52 controls underwent some oximetry. A change in medical management attributed to cerebral hypoxia was recorded in 222/772 intervention infants. Cardiovascular support within 72 hours occurred in 36.9% versus 30.2%. Cerebral oxygenation values themselves were not collected centrally, so achieved separation in cerebral hypoxia was unavailable.^{1,3}

Death or severe brain injury occurred in 272/772 infants (35.2%) versus 274/807 (34.0%): relative risk 1.03 (95% CI 0.90–1.18; P=0.64). Death occurred in 164/772 versus 160/807, relative risk 1.07 (95% CI 0.89–1.28). Severe brain injury was recorded in 185/763 versus 188/797 infants with relevant data, relative risk 1.02 (95% CI 0.85–1.21). This injury component included infants who subsequently died and should not be added to mortality to reconstruct the composite.¹

Exploratory composites showed no convincing benefit. Death or bronchopulmonary dysplasia occurred in 484/736 versus 524/772, relative risk 0.96 (95% CI 0.89–1.02); death or necrotising enterocolitis occurred in 232/772 versus 209/807, relative risk 1.16 (95% CI 1.00–1.35). At least one serious adverse event affected 657/772 versus 698/807, relative risk 0.98 (95% CI 0.94–1.01). Four serious adverse reactions, all severe skin injuries, occurred with cerebral oximetry. Primary sensitivity analyses, including adjustment for sibling dependence, did not materially alter the result. Two-year neurodevelopmental follow-up was ongoing at the time of this report.

Critique

The principal result is clear: implementing cerebral oximetry with suggested treatment responses did not reduce death or severe ultrasound-defined brain injury in this population. The trial reached its planned size, the control event rate matched the assumption used for power, and primary follow-up was nearly complete. The confidence interval excludes the large 22% relative benefit sought in the design. It still permits a smaller benefit and appreciable harm, so the result is not proof of equivalence. Nevertheless, it provides substantially stronger evidence against routine adoption for this indication than the earlier physiological trial provided in favour of it.

The largest explanatory gap is the absence of cerebral oxygenation recordings. SafeBoosC-III cannot establish whether the strategy reduced cerebral hypoxia, whether treatment normalised readings when attempted, or whether a specific device or subgroup achieved a meaningful physiological contrast. This prevents separation of

failure to implement the intervention from failure of the proposed mechanism to improve clinical outcomes. It does not invalidate the pragmatic answer: offering this strategy under the tested conditions failed to improve the primary endpoint. A hypothetical benefit under more effective implementation requires new evidence rather than an assumption that the monitor would work if used differently.

Adherence was primarily a measure of monitor availability, not of appropriate responses to actionable hypoxia. Only 28.8% of intervention infants had a recorded management change attributed to low cerebral oxygenation. This could reflect infrequent actionable events, effective existing care, reluctance to intervene or other factors that the available data cannot distinguish. Training was offered but not compulsory. At six months, mean certification rates across centres were 35% for nurses and 47% for neonatologists, below the programme's 70% aim. These figures make variation in implementation plausible, without proving that insufficient training caused the null result.³

The intervention also permitted different treatment choices in response to the same warning. Increasing oxygen, giving blood, altering ventilation and treating circulatory dysfunction have different indications and potential harms. The excess cardiovascular support in the oximetry group demonstrates that some care differed after allocation, but it is a post-randomisation observation. It cannot be treated as a baseline imbalance to adjust away, nor does it identify whether additional treatment was useful or unnecessary. Without physiological and detailed response data, the study cannot isolate which part of the monitoring-and-treatment pathway contributed to the final outcome.

Timing limits the biological claim. Monitoring generally started at three hours, after delivery-room transition, and stopped at 72 hours despite a neonatal course often lasting many weeks. Injury before monitoring began or arising later would not necessarily be preventable by this intervention. Moreover, the composite includes deaths and brain lesions from mechanisms that may not respond to correction of a low regional oxygenation value. The endpoint is clinically more meaningful than a saturation burden, but the proportion of its events actually modifiable through this pathway remains unknown. Neither earlier initiation nor a longer monitoring period was tested here.

Outcome ascertainment had both protections and limitations. Blinded brain-injury assessment reduced the risk that knowledge of treatment directly altered classification. However, local scanning and reports supplied the material being assessed. Both early and late scans were recorded in 76.2% of monitored infants and 79.1% of controls. A post hoc analysis excluding survivors without late scans gave similar results, but cannot recover lesions missed in the original assessment. Central adjudication and magnetic

resonance imaging might have provided different information. Ultrasound-defined injury at 36 weeks is also not synonymous with later cognitive, behavioural or motor disability.^{1,3}

Mortality and injury must be interpreted together without double counting. Death prevents subsequent assessment, while an infant can have diagnosed injury and then die. The primary composite deals with this more appropriately than an injury comparison restricted to survivors alone. The separate injury denominator is smaller than the mortality denominator because some infants had no usable scan information. Nine infants were recorded as never scanned despite having injury data; all died early, so the primary composite remained classifiable through mortality. That reported inconsistency is relevant to component ascertainment but does not by itself explain the null primary comparison.

The analysis contains useful safeguards. Independent masked analyses and agreement on alternative interpretations before unmasking reduce opportunities for outcome-driven interpretation. Accounting for sibling dependence is particularly relevant because multiple-birth infants shared allocation. The corresponding sensitivity estimate remained close to the primary result. Extreme assumptions about the 22 missing primary outcomes also failed to produce a large benefit. Conversely, the site meta-analysis's low heterogeneity does not prove every centre implemented the strategy equally well: many contributed few infants and had very imprecise estimates.

Reported amendments deserve proportionate attention. The per-protocol definition was relaxed from continuous monitoring to allowing less than 14 hours missing among infants surviving beyond 72 hours, and effect reporting changed from odds ratios to relative risks. The appendix places these decisions before analysis. The bronchopulmonary dysplasia definition was expanded to include nasal high flow early in recruitment. The death-inclusive exploratory outcomes were explicitly post hoc. These changes are not evidence of manipulation, but their chronology and status should remain visible. A contradictory control denominator in the supplementary per-protocol table also prevents precise reconstruction of that secondary analysis; its effect estimate should not displace the reliable primary comparison.³

Finally, absence of a significant aggregate safety difference is not proof that adding a sensor is harmless. Four severe skin injuries occurred, and management prompted by a monitor creates risks beyond the sensor itself. The upper confidence bound for the primary endpoint allows harm, while the exploratory necrotising-enterocolitis composite points unfavourably but remains insufficient for a definitive causal claim, especially with multiple comparisons. The evidence applies to infants selected for full intensive care

before 28 weeks' gestation. It does not determine the value of cerebral oximetry in a specific diagnostic episode, another neonatal population or a different treatment algorithm, and it leaves long-term outcomes unresolved within the 2023 evidence period.

Body of Evidence

The phase II SafeBoosC trial randomised 166 extremely preterm infants to visible cerebral oximetry with a treatment guideline or masked oximetry with standard care. Its primary measure combined duration and depth of readings outside 55–85%. Median burden was 36.1 versus 81.3 percentage-hours, a reported reduction of 58%, driven predominantly by less hypoxia. Clinical outcomes did not differ significantly. The crucial achievement was demonstration that the monitored physiological exposure could be changed. It justified a clinical endpoint trial but did not establish that reducing this particular surrogate would prevent severe injury or death.²

Follow-up of the same phase II cohort examined neurodevelopment at two years' corrected age. Among 115 assessed survivors, developmental and parental questionnaire scores did not differ convincingly between allocation groups. Moderate-to-severe impairment was recorded in 10 of 65 intervention children and six of 50 controls. This was a small follow-up comparison, not an independent confirmatory trial, and neither benefit nor harm was precisely excluded. Its relevance to SafeBoosC-III was the continued absence of demonstrated functional benefit despite the earlier improvement in cerebral oxygenation.⁴

COSGOD III, published in 2023, tested a different window: the first 15 minutes after birth in infants born before 32 weeks. Among 607 infants analysed, survival without cerebral injury occurred in 252/304 with cerebral oximetry-guided care and 238/303 with standard care, relative risk 1.06 (95% CI 0.98–1.14). Cerebral tissue oxygenation was similar between groups. The study therefore did not establish a clinical advantage from adding this strategy during immediate transition. Its broader gestational range, earlier intervention and injury definition prevent it from being treated as a replication of the SafeBoosC-III regimen.⁵

The COSGOD III interpretation was corrected during 2023 to remove claims that extrapolated uncertain numerical differences into substantial population benefit. That correction is relevant to how both trials should be read: an attractive point estimate does not establish an effect when uncertainty remains broad. Equally, SafeBoosC-III should not be described as proving that cerebral hypoxia is unimportant merely because

the tested strategy failed. These are treatment-policy trials, and each policy includes measurement, recognition, action and the consequences of those actions.⁶

The sequence of evidence is therefore informative without being contradictory. The phase II programme established feasibility and an effect on a physiological measure. Its follow-up did not demonstrate a long-term clinical gain. The much larger phase III trial then tested short-term survival and brain injury directly and found no benefit from the pragmatic strategy. Differences in training, devices, timing and response may help explain why a physiological result did not translate, but SafeBoosC-III did not collect the recordings needed to choose among those explanations. They remain research questions rather than established reasons to disregard the clinical endpoint.

By the end of 2023, routine cerebral oximetry for the first 72 hours in extremely preterm infants lacked evidence of improved survival without severe brain injury. Further evaluation would need to specify what clinicians should do with an abnormal reading, demonstrate that the resulting intervention changes the intended exposure and assess outcomes extending beyond neonatal imaging. Better physiological targeting might be valuable, but its value cannot be assumed from measurement capability alone. For current interpretation of this trial, the relevant decision is whether to add the tested strategy routinely; the evidence does not support doing so to achieve the claimed reduction in death or severe brain injury.

Should cerebral oximetry routinely guide care during the first 72 hours in extremely preterm infants?

SafeBoosC-III does not support routine use to reduce death or severe brain injury. Whether a different implementation improves long-term neurodevelopment remains unresolved.

References

1. Hansen ML, Pellicer A, Hyttel-Sørensen S, Ergenekon E, Szczapa T, Hagmann C, et al. Cerebral Oximetry Monitoring in Extremely Preterm Infants. *N Engl J Med*. 2023;388(16):1501–1511.
2. Hyttel-Sorensen S, Pellicer A, Alderliesten T, Austin T, van Bel F, Benders M, et al. Cerebral near infrared spectroscopy oximetry in extremely preterm infants: phase II randomised clinical trial. *BMJ*. 2015;350:g7635.

3. Hansen ML, Pellicer A, Hyttel-Sørensen S, et al. Supplementary appendix to: Cerebral Oximetry Monitoring in Extremely Preterm Infants. *N Engl J Med*. 2023;388:1501–1511.
4. Plomgaard AM, Alderliesten T, van Bel F, Benders M, Claris O, Cordeiro M, et al. No neurodevelopmental benefit of cerebral oximetry in the first randomised trial (SafeBoosC II) in preterm infants during the first days of life. *Acta Paediatr*. 2019;108(2):275–281.
5. Pichler G, Goeral K, Hammerl M, Perme T, Dempsey EM, Springer L, et al.; COSGOD III study group. Cerebral regional tissue Oxygen Saturation to Guide Oxygen Delivery in preterm neonates during immediate transition after birth (COSGOD III): multicentre randomised phase 3 clinical trial. *BMJ*. 2023;380:e072313.
6. Correction: Cerebral regional tissue Oxygen Saturation to Guide Oxygen Delivery in preterm neonates during immediate transition after birth (COSGOD III): multicentre randomised phase 3 clinical trial. *BMJ*. 2023;381:p1102.

PARIS-2

Franklin D, Babl FE, George S, Oakley E, Borland ML, Neutze J, et al. Effect of Early High-Flow Nasal Oxygen vs Standard Oxygen Therapy on Length of Hospital Stay in Hospitalized Children With Acute Hypoxemic Respiratory Failure: The PARIS-2 Randomized Clinical Trial. JAMA. 2023;329(3):224–234.

Introduction

High-flow nasal oxygen offers an appealing way to support a child with hypoxaemia and increased work of breathing. It can be delivered outside intensive care, supplies warmed humidified gas and provides a more substantial respiratory intervention than conventional low-flow oxygen. These practical advantages encouraged use early in acute respiratory illness, before non-invasive or invasive ventilation became necessary. The unresolved question was whether starting it routinely improved recovery sufficiently to justify the additional treatment.¹

Earlier paediatric trials had shown fewer episodes labelled treatment failure or escalation when high flow was used initially. However, avoiding escalation is not necessarily equivalent to shortening illness, preventing intensive care admission or reducing hospital stay. If a child begins on the more intensive treatment, there is also less opportunity to cross a treatment-escalation threshold. Studies permitting rescue high flow after conventional oxygen failure can therefore inform a staged management strategy as well as a comparison of the initial devices.²

Population matters. Much of the preceding evidence concerned infants with bronchiolitis, whereas preschool children have a broader mixture of wheezing disorders, pneumonia and other respiratory illnesses. A pilot trial extended the question beyond infant bronchiolitis and supported the feasibility of a larger comparison. It did not establish a reduction in hospital stay or intensive care admissions. That left uncertainty over whether routine early high flow helped children whose oxygen requirement persisted after initial emergency treatment but who did not yet require critical care.³

PARIS-2 addressed this clinical decision in children aged one to four years. Its primary hypothesis was that early high flow would shorten hospital stay compared with standard oxygen, with treatment changes available when clinically required. A useful result would need to distinguish the physiological capabilities of high flow from the consequences of introducing it into everyday care, including tolerance, weaning and decisions about escalation or discharge.

Synopsis

PARIS-2 was an investigator-initiated, multicentre, randomised superiority trial conducted in 14 Australian and New Zealand hospitals, comprising tertiary and regional services. Recruitment took place from 18 December 2017 to 18 March 2020, with final follow-up completed on 22 March. Public and charitable research organisations funded the study; Fisher & Paykel supplied equipment and consumables without charge. Funders and sponsors had no stated role in design, conduct, analysis, interpretation or publication decisions.¹

Children aged one to four years were eligible when an acute respiratory illness caused increased work of breathing, respiratory rate of at least 35/min, a continuing oxygen requirement and a need for hospital admission. Initial emergency management, including bronchodilators and other clinically indicated treatments, preceded enrolment. Hypoxaemia had to persist in room air below the local saturation threshold, generally 92%, with one hospital using 90%. The population included wheezing illness, pneumonia, pneumonitis and bronchiolitis in children older than 12 months.

Important exclusions were upper airway obstruction, craniofacial abnormalities, cyanotic heart disease and an immediate requirement for intensive care or ventilatory support. The flow diagram also identifies prior high flow during the episode, home oxygen and more than four hours of oxygen before screening as exclusion reasons. Parents provided prospective or deferred consent. Randomisation was generally computer-generated online in blocks of ten, stratified by hospital and the presence or absence of audible wheeze; one centre used opaque sealed envelopes. Clinical treatment could not be masked. Investigators remained unaware of aggregate trial outcomes until recruitment was complete and data were locked.¹

The high-flow group received AIRVO-2 treatment through an appropriate OptiFlow interface. Flow was 2 L/kg/min up to 12 kg, 30 L/min for 13–15 kg, 35 L/min for 16–30 kg and 40 L/min for 31–50 kg. Inspired oxygen was adjusted to a saturation of 92–98%, or 90–98% at the centre using the lower threshold. During bronchodilator administration, high flow was stopped and standard oxygen supplied. Flow itself was not progressively reduced: inspired oxygen was weaned to room air once the child was clinically stable, and high flow stopped when saturation remained acceptable at an inspired oxygen fraction of 0.21.

Standard treatment used subnasal cannulae up to 2 L/min or a Hudson mask up to 8 L/min, targeting the applicable saturation range. Oxygen was reduced once the child was stable. Switching to the alternative therapy was allowed for intolerance or clinical escalation. Recommended escalation thresholds included inspired oxygen above 0.40 or

0.50 during high flow, or conventional oxygen exceeding the specified nasal or mask limits. Treating teams retained responsibility for escalation and intensive care admission. This compared initial strategies with subsequent clinical adaptation, not compulsory uninterrupted use of either device.¹

The primary outcome was hospital length of stay from randomisation to discharge or death. Secondary outcomes included oxygen-treatment duration, total stay from emergency presentation, change of oxygen modality, intensive care admission or transfer, non-invasive or invasive ventilation, tolerance and adverse events. High-flow treatment duration extended until the device was stopped, including any period on room air. Serious adverse events included death, life-threatening events, disability or events prolonging hospitalisation. Economic outcomes were designated for separate reporting.

The investigators sought a reduction of 0.4 days, or 9.6 hours, from an assumed median stay of two days. A survival-analysis calculation required 1,209 children for 90% power with a two-sided 5% significance level; allowing 20% non-adherence increased the target to 1,512. The principal Cox model adjusted for wheeze category and included hospital as a shared frailty effect. A hazard ratio below one represented slower hospital discharge with high flow. Quantile regression also estimated the difference between group medians. No multiplicity correction was used for secondary outcomes, which therefore required exploratory interpretation.¹

The report identifies 23,862 screened children and 3,030 eligible children. A total of 1,567 were randomised, 782 to high flow and 785 to standard oxygen. Deferred consent was unavailable or refused for 29 and 21 children, leaving 753 and 764 in the primary analysis. One further consent withdrawal permitted retention of data. Hospital-stay information was complete among analysed children, and the planned missing-data imputation was not used.

Median age was 1.9 years in both groups, median weight approximately 12.5–12.7 kg and median room-air saturation 88%. Audible wheeze was present in about 70%. Approximately 55% had previously been admitted with respiratory disease. Randomisation occurred around three hours after emergency presentation. Before enrolment, 93.6% of randomised children had received bronchodilators, 63.4% steroids and 40.2% antibiotics. The clinical-management results state that 714/753 and 735/764 initially received their allocated treatment, rather than universal initial adherence.¹

Hospital stay was longer with high flow: median 1.77 days (IQR 1.03–2.80) versus 1.50 days (IQR 0.85–2.44), adjusted hazard ratio 0.83 (95% CI 0.75–0.92; $P < 0.001$). The reported median difference was 0.27 days (95% CI 0.13–0.41), approximately 6.5 hours

when independently converted from days. Total stay from emergency presentation was also longer, with an adjusted hazard ratio of 0.82 (95% CI 0.74–0.91). Oxygen-treatment duration was 1.07 versus 0.75 days, adjusted hazard ratio 0.78 (95% CI 0.70–0.86).

Intensive care admission occurred in 94/753 children (12.5%) versus 53/764 (6.9%), adjusted odds ratio 1.93 (95% CI 1.35–2.75), with a reported absolute difference of 5.5 percentage points. Requiring additional clinical escalation criteria in a sensitivity analysis yielded 58 versus 33 events and an adjusted odds ratio of 1.87 (95% CI 1.20–2.90). Seven versus six children required non-invasive ventilation and four versus none invasive ventilation. These were small event counts; the trial did not establish a mortality effect.^{1,4}

Switching to the alternative oxygen therapy occurred in 323/753 versus 141/764, adjusted odds ratio 3.42 (95% CI 2.70–4.34). Intolerance was recorded as a reason in 92 high-flow children and no standard-oxygen children; increased work of breathing was recorded in 33 versus 98. Reasons could overlap. The per-protocol cohort comprised 336 versus 498 children and retained a longer stay with high flow. There was no convincing treatment interaction by wheeze or age. One death occurred in the high-flow group and was judged unrelated to treatment. No air leak, emergency intubation, cardiac arrest or respiratory arrest was reported; two high-flow children had nosebleeds.^{1,4}

Critique

PARIS-2 does more than fail to show that routine high flow shortens admission. It demonstrates a statistically supported prolongation of hospital stay under the tested strategy, with longer oxygen treatment and more intensive care use pointing in the same direction. Describing it solely as a negative superiority trial obscures this finding. The median prolongation was modest for an individual child, but additional hours of treatment and admission matter when a therapy is applied broadly to a common condition. The evidence therefore argues against routinely beginning high flow in every comparable child after initial emergency management.

That conclusion concerns a care strategy, not proof that high flow worsens pulmonary injury. Hospital stay is determined partly by physiological recovery and partly by how clinicians respond to treatment. A child receiving a more conspicuous support device may be perceived as needing continued monitoring or more cautious weaning. Such effects are relevant to pragmatic effectiveness, even if the underlying lung disease is unchanged. The trial's inability to mask clinicians makes it difficult to disentangle these pathways, but does not make the additional resource use disappear.

Weaning is particularly important because it lies directly between the assigned intervention and discharge. High-flow rates remained fixed until cessation, with inspired oxygen reduced first. Time on high flow at an oxygen fraction of 0.21 still counted towards treatment duration. The corresponding standard-oxygen measure therefore need not represent exactly the same mixture of oxygen exposure and respiratory support. Detailed weaning data were not collected, so it is unknown whether prolonged support reflected continuing physiological need, reluctance to stop the device or another process. The common saturation target does not ensure that withdrawing the two modalities was behaviourally identical.¹

The primary outcome was appropriately analysed from randomisation, protecting the comparison from differences in pre-enrolment waiting time. Results measured from emergency presentation were consistent. The adjusted hazard ratio describes the rate of leaving hospital over time; it is not a 17% increase in length of stay. The quantile estimate supplies a more interpretable absolute comparison but does not mean every child stayed 6.5 hours longer. Death also ended the length-of-stay measure. With one death this is unlikely to dominate the overall result, but the endpoint should not be described as survival to successful discharge in every participant.

Frequent crossover is central to the clinical question. More than 40% of high-flow children subsequently received standard oxygen, compared with fewer than 20% moving in the opposite direction. A switch away from poorly tolerated high flow is different from adding high flow because conventional treatment is insufficient. The pre-switch data support that distinction: children moving from standard oxygen had higher respiratory rates and early-warning scores, and more often had moderate or severe work of breathing. These are selected post-randomisation observations, not proof of a causal physiological advantage for either therapy.⁴

The allocation-based comparison remains the most useful estimate of what happens when a hospital starts with one strategy and permits rescue or de-escalation. In contrast, removing children who switch selects on clinical deterioration, response and tolerance. The per-protocol cohort also required the eligibility criteria to be met, explaining why it was smaller than the simple non-crossover totals. Its similar length-of-stay direction is supportive but does not prove crossover was irrelevant. The striking intensive care imbalance in that selected cohort is especially vulnerable to excluding standard-oxygen children who deteriorated and received rescue high flow. It should not replace the randomised estimate.

Intensive care admission was a secondary and partly discretionary endpoint. Clinicians could lower their transfer threshold because a child was already receiving perceived

higher-level support, while local ward capacity could influence where treatment continued. The sensitivity analysis requiring additional deterioration criteria helps address purely discretionary admission and retains an unfavourable estimate. Nevertheless, it does not establish that high flow caused more severe respiratory failure. Non-invasive and invasive ventilation were uncommon, with too few events for a reliable mechanism-specific conclusion. The observed excess in intensive care use is clinically important without being equivalent to an established excess in intubation or death.

Tolerance reporting contains a substantive discrepancy. The main narrative describes no significant difference in comfort scores, but the supplement's first parental and staff ratings have confidence intervals excluding zero in the direction of greater discomfort with high flow. Later change scores do not show a clear difference. Incomplete sampling and measurements away from the moment of crossover further limit interpretation. It is therefore inappropriate to dismiss intolerance on the basis of the main narrative alone; the 92 treatment switches explicitly attributed to intolerance remain relevant.^{1,4}

Randomisation, multicentre recruitment and complete primary follow-up among analysed children are concrete strengths. However, 50 children were excluded after allocation because consent was unavailable, so "no missing length-of-stay data" applies to the analysed population rather than every randomised child. The exclusions were relatively few but were not outcome observations. The analysed denominators, primary medians and effect estimate are consistently reported. These strengths support interpretation of the primary comparison, while the exclusions limit its scope.

Generalisability is strongest to preschool children with mild-to-moderate hypoxaemic respiratory failure who remain suitable for ward care. Immediate critical care or ventilatory needs were excluded. Most had wheeze, but the bedside wheeze classification was not the same as a final diagnosis of asthma or reactive airway disease. The absence of a convincing interaction does not identify a subgroup in whom routine high flow is beneficial. Infants younger than one year, older children and settings with different staffing or escalation capacity require separate evidence. PARIS-2 addresses routine initial high flow; it does not establish that rescue high flow should never be used after standard oxygen proves inadequate.

Body of Evidence

The 2017 HFWHO trial enrolled 202 children younger than two years with moderate bronchiolitis. Initial high flow reduced treatment failure, but did not shorten the primary outcome of time to final oxygen cessation: median 20 versus 24 hours, without a significant difference. Twenty of 33 children who failed standard treatment responded

to rescue high flow. This separated a useful rescue response from proof that starting every child on high flow accelerated recovery. Its lower flow regimen and younger bronchiolitis population also differed from PARIS-2.⁵

The larger 2018 bronchiolitis trial analysed 1,472 infants younger than 12 months treated outside intensive care. Escalation for treatment failure occurred in 87/739 receiving high flow and 167/733 receiving standard oxygen, an absolute reduction of 11 percentage points. Hospital-stay and oxygen-duration differences were not significant. Of the standard-treatment failures, 102/167 responded to rescue high flow. The trial supported reduced escalation under its definitions, but did not show that early high flow shortened admission. It also supplied evidence that a conventional-first strategy could retain high flow as a subsequent option.²

The PARIS pilot broadened recruitment to children aged up to 16 years in two tertiary hospitals, randomising 563 participants. It suggested fewer treatment failures with high flow but did not demonstrate shorter admission or reduced intensive care use. Approximately 60% of standard-oxygen failures responded to rescue high flow. A more favourable result in children with obstructive disease was a reason for further study, not an established subgroup indication. PARIS-2 subsequently provided a larger comparison in the age range accounting for most pilot participants and chose hospital stay as its principal endpoint.³

These studies addressed related but different questions. An escalation endpoint is useful when it identifies children who need additional care, yet it depends on the starting treatment and the available rescue pathway. Beginning with high flow can reduce the need to move from standard oxygen to high flow without necessarily preventing intensive care or changing the duration of illness. PARIS-2's choice of hospital stay tested whether a broader patient and service benefit followed from routine early treatment. Its unfavourable result therefore does not contradict the earlier observations simply because their escalation outcomes favoured high flow.

The comparison also illustrates why clinical labels need care. "Acute hypoxaemic respiratory failure" spans infants with bronchiolitis, preschool wheezing illness and children requiring immediate ventilation. Treatments may perform differently across those settings, and both baseline severity and the comparator's rescue options influence the effect. PARIS-2 included persistent hypoxaemia after initial treatment but excluded the sickest children. Its findings should neither be diluted by evidence from a different age group nor expanded into a universal statement about all paediatric respiratory support.

By the end of 2023, the evidence supported selective use of high flow more strongly than universal early application in preschool children. PARIS-2 showed that introducing the device routinely could prolong treatment and admission despite its physiological appeal. The trial left uncertainty over the contribution of tolerance, weaning behaviour and underlying recovery, but the practical outcome was clear enough to inform the initial strategy. Standard oxygen with reassessment and access to escalation remained a reasonable approach for children matching the trial population. Future improvements would need to show that a different selection or withdrawal strategy actually changes patient outcomes, rather than assuming that a more intensive device necessarily provides better care.

Should high flow replace standard oxygen as routine initial treatment in preschool children?

No. In children matching PARIS-2 eligibility, routine early high flow prolonged hospital stay and increased ICU use; the trial does not exclude a selective rescue role.

References

1. Franklin D, Babl FE, George S, Oakley E, Borland ML, Neutze J, et al. Effect of Early High-Flow Nasal Oxygen vs Standard Oxygen Therapy on Length of Hospital Stay in Hospitalized Children With Acute Hypoxemic Respiratory Failure: The PARIS-2 Randomized Clinical Trial. *JAMA*. 2023;329(3):224–234.
2. Franklin D, Babl FE, Schlapbach LJ, Oakley E, Craig S, Neutze J, et al. A Randomized Trial of High-Flow Oxygen Therapy in Infants with Bronchiolitis. *N Engl J Med*. 2018;378(12):1121–1131.
3. Franklin D, Shellshear D, Babl FE, Hendrickson R, Williams A, Gibbons K, et al.; PARIS and PREDICT. High flow in children with respiratory failure: A randomised controlled pilot trial – A paediatric acute respiratory intervention study. *J Paediatr Child Health*. 2021;57(2):273–281.
4. Franklin D, Babl FE, George S, et al. Supplement 3 to: Effect of Early High-Flow Nasal Oxygen vs Standard Oxygen Therapy on Length of Hospital Stay in Hospitalized Children With Acute Hypoxemic Respiratory Failure: The PARIS-2 Randomized Clinical Trial. *JAMA*. 2023.

5. Kepreotes E, Whitehead B, Attia J, Oldmeadow C, Collison A, Searles A, et al. High-flow warm humidified oxygen versus standard low-flow nasal cannula oxygen for moderate bronchiolitis (HFWHO RCT): an open, phase 4, randomised controlled trial. *Lancet*. 2017;389(10072):930–939.

Possible Reporting Inconsistencies. The Results and Discussion report no significant comfort difference, but supplementary eTable 3 gives initial adjusted differences of 0.88 (95% CI 0.46–1.31) for parent ratings and 0.99 (0.59–1.39) for staff ratings, both indicating greater discomfort with high-flow therapy. These intervals exclude zero. This reporting conflict concerns the initial ratings and does not establish a clinically important or sustained discomfort difference.^{1,4}



Miscellaneous Trials

Early Mobilisation in ICU

Patel BK, Wolfe KS, Patel SB, Dugan KC, Esbrook CL, Pawlik AJ, et al. Effect of early mobilisation on long-term cognitive impairment in critical illness in the USA: a randomised controlled trial. *Lancet Respir Med*. 2023;11(6):563–572.

Introduction

Leaving intensive care does not necessarily restore a patient's ability to think, organise everyday activities or live independently. Cognitive recovery is therefore a distinct therapeutic objective alongside survival and physical rehabilitation. The BRAIN-ICU cohort linked longer delirium duration with worse subsequent cognition, but an association between an acute syndrome and later disability does not establish that shortening delirium will prevent that disability. An intervention needs to demonstrate a benefit beyond an encouraging change during the admission.¹

Early mobilisation is a plausible candidate because meaningful activity combines movement with attention, interaction and the performance of familiar tasks. Physical and occupational therapists can coordinate treatment with periods of wakefulness, progressing from simple movement to sitting, standing and activities of daily living. This is a complex intervention: its possible effects include earlier activity, greater human interaction, functional practice and changes in how the clinical team organises care. A programme described simply as exercise may not reproduce the same treatment.²

A 2009 randomised trial found better functional independence at hospital discharge when early physical and occupational therapy accompanied daily sedation interruption. It also found less delirium. Those short-term findings left open whether cognition improved after recovery, whether effects persisted, and whether more recent usual care left sufficient scope for further benefit. Delivering therapy safely during invasive ventilation also required attention to physiological instability and device security.³

The University of Chicago trial addressed cognition one year after hospital discharge in previously functionally independent adults receiving mechanical ventilation. It compared coordinated early physical and occupational therapy with therapy initiated by the treating team or after extubation. The hypothesis was that intervention during the first days of critical illness would prevent enduring cognitive impairment. Its interpretation requires equal attention to the treatment actually delivered and to how cognitive outcomes were handled when patients died or could not be tested.

Synopsis

This was a single-centre, parallel-group randomised trial in an adult medical-surgical ICU at the University of Chicago. Recruitment ran from 11 August 2011 to 24 October 2019. The report states that there was no study funder, while acknowledging external career and training support for investigators. Ethics approval preceded recruitment, but trial registration occurred in January 2013. The investigators attributed the delay to mistakenly believing that an earlier mobilisation trial's registration covered this study and stated that no preliminary data were evaluated before the new registration.²

Eligible patients were at least 18 years old, had a premorbid Barthel score above 70, had received mechanical ventilation for less than 96 hours and were expected to require at least another 24 hours. Exclusions included cardiac arrest, rapidly changing neurological disease, raised intracranial pressure, traumatic brain injury, a terminal condition with life expectancy below six months, pregnancy and important conditions preventing mobilisation, including multiple limb or pelvic fractures. Surrogates provided written consent. Computer-generated randomisation used varying balanced blocks and a 1:1 ratio. A separate investigator placed assignments in consecutively numbered opaque sealed envelopes and had no subsequent trial involvement. Follow-up assessors were masked, had no access to patient records and took no part in other trial activities; treatment teams could not be masked.²

Intervention patients received joint physical and occupational therapy after sedation interruption. Activities progressed according to tolerance from range-of-motion exercises and bed mobility to upright positioning, sitting, standing, marching and walking. Cueing encouraged interaction and command following, and seated activities incorporated everyday tasks. Daily sessions lasted approximately 25–30 minutes and continued during hospitalisation until discharge or return to baseline function. Therapists were experienced in critical care, and different therapists treated the two groups. Usual-care therapy began when ordered by the primary team or on extubation, whichever came first.²

Both groups received goal-directed sedation, daily sedation interruption and paired awakening and spontaneous breathing trials. Early therapy was therefore added to an existing approach to sedation and ventilator liberation. Reasons to withhold or stop therapy included mean arterial pressure below 65 or above 110 mmHg, systolic pressure above 200 mmHg, heart rate below 40 or above 130/min, respiratory rate below five or above 40/min and oxygen saturation below 88%. New arrhythmia, ischaemia, marked ventilator asynchrony, distress, unsecured airway, active bleeding and other specified instability also precluded treatment. These were safety conditions for tailored

progression rather than a requirement to reach a fixed mobility level regardless of physiology.²

The primary endpoint was cognitive impairment one year after hospital discharge, defined by a Montreal Cognitive Assessment (MoCA) score below 26 out of 30. Masked assessments occurred at hospital discharge and at one year, with home visits when necessary. Secondary outcomes included discharge cognition, weakness defined by a Medical Research Council strength score below 48, functional independence, SF-36 quality of life and days alive at home during the following year. Functional independence required activities of daily living without physical assistance and independent ambulation. Deaths were also assessed. Adverse events were observed in real time in the intervention group but identified from therapy records in usual care.²

The sample-size calculation sought a reduction in cognitive impairment from 50% to 30%. A longitudinal difference-in-proportions calculation, assuming an intraclass correlation of 0.6, required 75 participants per group for 80% power at a 5% significance threshold. Allowance for anticipated in-hospital and subsequent mortality increased recruitment to 200. Categorical comparisons used chi-squared or Fisher's exact tests and continuous comparisons used rank-sum tests. A modified Poisson generalised estimating equation model examined repeated cognitive impairment; pattern-mixture modelling explored incomplete follow-up. A cause-specific hazards analysis treated death as a competing event for the first test showing cognitive impairment. Survival analysis was post hoc. Reported P values were two-sided without multiplicity adjustment.²

Of 1,222 screened patients, 200 were randomised, 100 per group. One participant withdrew from each arm, leaving 99 per group in the reported intention-to-treat analysis. Median age was 57.9 years with early mobilisation and 54.5 with usual care; median APACHE II score was 23 in both groups and median premorbid Barthel score was 100. Most participants were African American, and approximately nine in ten had completed high school. Respiratory failure predominated, with sepsis present in 64% and 57%, respectively. This was a selected population with substantial acute illness but good preceding functional status.²

The achieved timing contrast was large. First therapy occurred a median 1.1 days after intubation with early mobilisation versus 4.7 days with usual care. Ninety-three of 99 intervention patients received therapy during mechanical ventilation within 96 hours; only six usual-care patients received therapy while ventilated. Median session counts during ventilation were two versus zero, and during hospitalisation five versus two. Median time to standing was 1.3 versus five days and to walking 1.7 versus 5.6 days.

Median delirium duration was zero versus one ICU day. Duration of mechanical ventilation, ventilator-free days and ICU or hospital stay did not differ significantly.^{2,4}

At one year, 70 intervention and 74 usual-care participants were alive. The report describes completed cognitive testing in 62 and 65, respectively, although its accompanying description of patients unable or unwilling to complete testing is inconsistent with those labels. Six patients in each arm were lost to follow-up. Using its full analysed-arm denominators, the principal report gives cognitive impairment in 24/99 intervention patients versus 43/99 usual-care patients: absolute difference –19.2 percentage points, 95% CI –32.1 to –6.3; $P=0.0043$. These are the published figures, not impairment rates among all surviving or all tested patients. Median one-year MoCA scores were 26 versus 23, with a reported difference of three points (95% CI 1–4; $P=0.0001$).²

At discharge, cognitive impairment was reported in 53/99 versus 68/99, an absolute difference of –15.2 percentage points (95% CI –28.6 to –1.7). The repeated-measures model gave a treatment risk ratio of 0.81 (95% CI 0.68–0.96), which is a different estimate from the single one-year contrast. The competing-risk analysis gave an APACHE II-adjusted cause-specific hazard ratio of 0.69 (95% CI 0.49–0.98). Pattern-mixture analysis retained a favourable difference in MoCA scores.^{2,4}

At one year, weakness was reported in 0/99 versus 14/99, and median SF-36 physical component scores were 52.4 versus 41.1. Functional independence, however, was similar: 64/99 versus 61/99, difference 3.0 percentage points (95% CI –10.4 to 16.5). Mental quality-of-life scores and institution-free days also showed no significant difference. Deaths by one year numbered 29 versus 25; the post-hoc survival comparison was not significant. The main article records seven adverse events in six intervention patients versus none in usual care, including haemodynamic changes, respiratory distress and device dislodgement. The supplementary adverse-event table reverses the arm labels, an unresolved inconsistency; the main article and abstract provide the basis for the chapter's safety figures.^{2,4}

Critique

The trial offers a clinically important signal that a coordinated early rehabilitation programme might improve later cognition. Its primary comparison was statistically convincing within the reported analysis, and the continuous MoCA difference points in the same direction as the binary threshold. However, the most important interpretative problem is what those primary percentages represent. Cognitive impairment is reported against 99 patients in each group despite only 70 and 74 surviving to one year and still

fewer completing assessment. A reader cannot interpret the complement of 24% or 43% as the proportion alive with normal cognition. Death and unobserved cognition are not evidence of cognitive recovery.²

The supplied reports do not clearly explain the contribution of every untested participant to that binary comparison. The flow diagram labels 62 and 65 as completing MoCA, while the text places two refusals and three patients unable to complete testing within similarly described denominators. This is more than a cosmetic reporting issue because it affects the interpretation of the endpoint and the confidence interval. It would be inappropriate to invent an imputation rule, silently replace the denominator with the number tested, or describe the result as a reduction in death or cognitive impairment. Clarification of the endpoint coding is needed before treating the absolute difference as a readily transportable treatment benefit.²

The sensitivity analyses are helpful but answer related questions. A cause-specific hazard ratio for the first impaired test is not a one-year probability of being alive without impairment. Pattern-mixture modelling of repeated MoCA scores explores differences associated with incomplete follow-up but cannot demonstrate that all plausible mechanisms of missingness would leave the result unchanged. In the appendix, the three-way dropout interaction is omitted because of collinearity. The continuous-score models retain a treatment difference without showing a significant treatment-by-time interaction, whereas the binary repeated-measures model reports an interaction. These are different outcomes and models; they should not be combined into a single claim that cognitive recovery accelerated across every analysis.^{2,4}

Nevertheless, this was not a trial in which assignment failed to change care. Therapy began several days earlier, most intervention patients received it while ventilated, and standing and walking occurred earlier. This strengthens the inference that the strategy created a meaningful contrast. The contrast was also broader than the amount of exercise. It combined physical and occupational therapy, coordinated access during wakefulness, functional tasks, additional attention and research-supported organisation. The trial cannot identify the independent contribution of occupational therapy, earlier timing or interpersonal engagement. Nor can it establish that simply increasing physiotherapy minutes would recreate the effect.²

The comparator is especially important for applicability. Both arms had daily awakening and breathing trials, yet usual-care mobilisation during mechanical ventilation was uncommon. The investigators describe coordination by research personnel in the intervention group that was absent in usual care. That is part of the tested service intervention and may be clinically valuable. It also means that the result most directly

informs a unit where therapy is commonly deferred until extubation or a separate referral. A unit already delivering early coordinated activity may have less room for improvement. The study does not quantify the marginal benefit over an established, consistently delivered early programme.²

A lower MoCA score is a meaningful signal requiring attention, but a screening threshold is not a complete diagnosis of disabling cognitive disease. Premorbid cognition was unknown: a Barthel score demonstrating physical independence does not exclude prior cognitive impairment. Randomisation provides protection against systematic baseline differences, although a study of this size cannot rule out chance imbalance. Masked assessors and home visits strengthen outcome collection, but masking success was not measured. Comprehensive neuropsychological testing and consequences such as return to work would help establish the practical meaning of the three-point score difference.²

The secondary findings are mixed rather than uniformly favourable. Less weakness and better physical quality of life support potential recovery benefits, but independence at one year, mental quality of life and time at home did not clearly improve. These results do not refute a cognitive effect: physical independence and mental health are different constructs. They do limit a broader claim that the programme prevents the whole syndrome of disability after intensive care. The many unadjusted secondary comparisons should remain supportive or exploratory, especially when follow-up completeness differs between instruments.^{2,4}

Safety deserves more than a statement that events were uncommon per session. Seven events occurred in six intervention patients, a clinically relevant patient-level experience. More treatment creates more opportunities for an event, while real-time surveillance detects problems that routine chart review can miss. These differences prevent a clean comparison of intrinsic risk per treatment session. They do not justify ignoring the events or calling the programme proven safe. Mortality was numerically higher with intervention, but the small trial neither establishes a mortality hazard nor excludes one. Consistent prospective surveillance in both arms is necessary in replication.²

The delayed registration and eight-year recruitment period add uncertainty. The authors explicitly state that preliminary data were not examined before registration, so delayed registration is not evidence of outcome selection after seeing results. It nonetheless weakens independent confirmation of the original commitments. Over a long recruitment period, staffing and routine care may change; the report does not establish how stable the treatment contrast was across calendar time. Combined with one experienced centre, exclusion of major neurological illness and a requirement for prior

independence, these features favour confirmation in a broader population before a universal claim of cognitive protection.²

Body of Evidence

The 2009 Schweickert trial randomised 104 functionally independent adults early during mechanical ventilation. Physical and occupational therapy during sedation interruption increased independence at discharge to 29/49 patients versus 19/55, with an odds ratio of 2.7 (95% CI 1.2–6.1). Delirium duration was shorter, and one serious desaturation event occurred during treatment. It supplied a rationale for coordinated early rehabilitation and a signal of short-term benefit. It did not establish one-year cognition, making the Chicago study an extension of the clinical question rather than a repeat of an already proven long-term treatment.³

BRAIN-ICU subsequently followed 821 adults with respiratory failure or shock. Longer delirium was independently associated with worse later global cognition and executive function, and deficits persisted among both younger and older survivors. This observational evidence established the importance of cognition as an outcome but did not randomise delirium prevention or rehabilitation. The Chicago trial's low delirium burden and shorter intervention-group delirium are compatible with several mechanisms; they do not demonstrate that delirium reduction mediated its cognitive result. Associations between delirium and cognition cannot identify which component of a multicomponent treatment needs to change.¹

In 2016, Morris and colleagues randomised 300 adults with acute respiratory failure to standardised rehabilitation or usual care. The intervention combined range of motion, physical therapy and progressive resistance exercise. Hospital stay was ten days in both groups, with no significant primary difference. Some physical-function measures were better at six months, but the Mini-Mental State Examination difference was not significant. That comparison illustrates both the variable content of rehabilitation programmes and the need to distinguish their primary outcome from selected later measures. It neither proves that all rehabilitation is ineffective nor independently confirms the Chicago cognitive result.⁵

TEAM provided the largest relevant randomised comparison available by the end of 2023. It assigned 750 ventilated adults to increased early mobilisation or usual care. Mean daily active mobilisation was 20.8 versus 8.8 minutes. Days alive and out of hospital at 180 days were 143 versus 145, difference –2 days (95% CI –10 to 6; P=0.62). Cognition and other survivor outcomes were similar. Potential mobilisation-related adverse events

occurred in 34/371 versus 15/370 patients. Thus a larger trial demonstrated treatment separation without better overall recovery and with more adverse events.⁶

The Chicago investigators propose timing and occupational therapy as explanations for disagreement with TEAM. Those are hypotheses worth testing, but their supplementary comparison also needs care: time to sitting or standing is a mobility milestone, not necessarily the first therapeutic contact. Cross-trial differences in patients, comparator treatment, sedation and follow-up are intertwined. They cannot isolate one causal explanation. In particular, an unfavourable or neutral larger trial should not be dismissed because its programme differed from the smaller positive study; the differences define the limits of what can be generalised.^{2,4}

Taken together, the evidence supports treating rehabilitation as a specified clinical programme rather than a uniform exposure labelled “early mobilisation”. Timing, tasks, the ability to participate, coordination with sedation and the comparator’s existing activity all shape the question being tested. The Chicago findings justify further investigation of early combined physical and occupational therapy, with transparent handling of death and missing cognition, comprehensive follow-up and equal adverse-event surveillance. They do not support a mandate to maximise activity despite instability, or an assurance that early mobilisation reliably prevents cognitive disability in every ventilated patient. By the book’s cut-off, the promising cognitive signal remained a finding requiring replication and clarification.

Does early combined physical and occupational therapy prevent long-term cognitive impairment?

The Chicago trial found a promising benefit, but incomplete outcome accounting and safety uncertainty require clarification and replication before cognitive protection can be considered established.

References

1. Pandharipande PP, Girard TD, Jackson JC, Morandi A, Thompson JL, Pun BT, et al.; BRAIN-ICU Study Investigators. Long-term cognitive impairment after critical illness. *N Engl J Med*. 2013;369(14):1306–1316.
2. Patel BK, Wolfe KS, Patel SB, Dugan KC, Esbrook CL, Pawlik AJ, et al. Effect of early mobilisation on long-term cognitive impairment in critical illness in the USA: a randomised controlled trial. *Lancet Respir Med*. 2023;11(6):563–572.

3. 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.
4. Patel BK, Wolfe KS, Patel SB, et al. Supplementary appendix to: Effect of early mobilisation on long-term cognitive impairment in critical illness in the USA: a randomised controlled trial. *Lancet Respir Med*. 2023.
5. Morris PE, Berry MJ, Files DC, Thompson JC, Hauser J, Flores L, et al. Standardized Rehabilitation and Hospital Length of Stay Among Patients With Acute Respiratory Failure: A Randomized Clinical Trial. *JAMA*. 2016;315(24):2694–2702.
6. Hodgson CL, Bailey M, Bellomo R, Brickell K, Broadley T, Buhr H, et al.; 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.

Possible Reporting Inconsistencies. Supplementary Table S4 appears to reverse the treatment arms for adverse events: it places the seven events in usual care, whereas the main abstract, Results and Table 4 place them in six intervention patients and report none with usual care. The main report’s allocation is used in this chapter.^{2,4}

PICSI-H

Curtis JR, Lee RY, Brumback LC, Kross EK, Downey L, Torrence J, et al. Intervention to Promote Communication About Goals of Care for Hospitalized Patients With Serious Illness: A Randomized Clinical Trial. JAMA. 2023;329(23):2028–2037.

Introduction

For a seriously ill patient admitted to hospital, decisions about treatment depend on more than the reversibility of the presenting illness. Clinicians also need to understand what the patient hopes treatment will achieve, which abilities matter most, and what burdens they would accept if recovery becomes uncertain. These questions extend beyond whether cardiopulmonary resuscitation should be attempted. An admission provides an opportunity to discuss them before deterioration forces decisions under pressure, but the opportunity can pass without either a conversation or an accessible record of its content. PICSI-H addressed that practical gap.¹

A prompt might help clinicians recognise the need for a discussion, find existing information and choose language with which to begin. Earlier Jumpstart trials had tested guides delivered to both clinicians and patients or their representatives, with encouraging effects on communication outcomes. However, obtaining patient information through study surveys required staff time and engagement before the intervention could be delivered. Whether a simpler approach could retain useful effects across a broader hospital population remained uncertain.^{2,3}

The electronic health record offers a possible way to identify eligible patients and assemble relevant information without a preliminary survey. It also provides an outcome that can be measured without contacting every patient after discharge. These advantages create a corresponding interpretative challenge: an intervention that encourages documentation might increase recorded discussions even when the quality, timing or patient experience of communication changes little. A usable record is valuable, but it is only one part of care that reflects a patient's goals.

PICSI-H therefore tested a deliberately limited intervention: a patient-specific, one-page guide sent to the hospital clinical team, containing existing record information and suggested conversation prompts. The hypothesis was that this clinician-facing approach would increase documented goals-of-care discussions compared with usual care. The trial also examined subsequent healthcare use and whether the effect differed across patient groups, including people from racially or ethnically minoritised populations.¹

Synopsis

This pragmatic, parallel-group randomised trial enrolled patients at three UW Medicine hospitals in the United States between April 2020 and March 2021. The sites comprised a university hospital, a county safety-net hospital and a community hospital within one health system. The intervention was intended to promote usual standards of communication, and the ethics committee authorised a waiver of informed consent. Clinicians could opt out of participation or request exclusion of an individual patient. Funding came from the National Institute on Aging and the Cambia Health Foundation, which had no reported role in conducting, analysing or publishing the study.¹

Patients were eligible after at least 12 hours but no more than 96 hours in hospital if they were aged at least 55 years and had a qualifying chronic condition identified from diagnostic codes recorded during the preceding two years. The nine categories included dementia, poor-prognosis cancer, chronic pulmonary disease, coronary artery disease, heart failure, chronic liver or kidney disease, diabetes with end-organ damage and peripheral vascular disease. All hospitalised patients aged at least 80 years were eligible irrespective of those conditions. Daily manual screening identified patients without a documented goals-of-care discussion or palliative care consultation during the admission. Important exclusions included recent transplantation, hospice or comfort care, psychiatric circumstances specified in the report, and protected records designated as having no available information. This was a hospital population, rather than an exclusively intensive care cohort.¹

Individual patients were allocated equally using variable blocks of two to six, stratified by hospital and history of dementia. Clinicians necessarily knew when they received the guide. The researchers abstracting the primary outcome were blinded to allocation, with disagreements resolved by consensus. The manuscript does not provide enough detail to independently establish how the next allocation was concealed before assignment.¹

The Jumpstart Guide combined automatically retrieved information on current resuscitation status, prior advance directives and physician orders for life-sustaining treatment with suggested approaches to a conversation. Example prompts concerned the patient's understanding of illness, abilities they would find difficult to live without, and priorities if illness worsened. Clinicians were encouraged to record a short account afterwards. The guide did not specify a treatment limitation or require a decision against resuscitation. It was sent once, by secure email on the day of randomisation, to the attending physician, resident physicians and advanced practice clinicians on the primary hospital team. A single pager message alerted recipients to the email. Control patients received usual care, including the information and documentation opportunities

ordinarily available in the record, without the study guide or alert. Neither arm received a study intervention directly addressed to patients or families.¹

The primary endpoint was a documented goals-of-care discussion within 30 days of randomisation. Qualifying discussions addressed overarching aims of medical care; documentation confined to resuscitation status did not count. Ascertainment used natural language processing to identify candidate passages in clinical notes, followed by human review. In an internal validation sample of 159 patients, this combined process had a reported patient-level sensitivity of 92.6%. Secondary outcomes included intensive care and emergency department use, palliative care consultation, death and time alive outside hospital or intensive care over 30 days, together with readmission within seven days of hospital discharge.¹

The planned sample was 2,000 patients, providing 80% power at a two-sided significance level of 0.05 to detect an absolute difference of six percentage points in the primary outcome. The calculation assumed an average event proportion of 0.54 across participants, rather than specifying that figure as the control event rate. Recruitment continued beyond 2,000 to increase the number with dementia and ended before a change in the health system's record vendor. Analyses followed intention to treat. A linear regression model with robust standard errors estimated the difference in outcome proportions, adjusting for hospital and dementia history. Effect modification was examined by age, sex, dementia, hospital, and minoritised race or ethnicity.¹

Of 3,918 patients assessed, 2,512 were randomised: 1,255 to intervention and 1,257 to usual care. Major reasons for exclusion included an existing discussion, discharge before screening, and transplantation or related treatments. All randomised patients were included in the primary analysis, with no reported loss to record-based follow-up. Median age was 70 years in each group, 42% were women, and 280 patients had dementia identified at allocation. Among 2,444 patients whose combined race and ethnicity classification was available, 803 belonged to the trial's minoritised category. This combined people identified as Hispanic or as a race other than White. Baseline advance directives were present in 7.7% versus 10.7%. The flow diagram reports delivery as allocated, but the publication does not quantify how often clinicians opened, read or used the guide.¹

The primary outcome occurred in 433 of 1,255 intervention patients (34.5%) and 382 of 1,257 controls (30.4%). The adjusted absolute difference was 4.1 percentage points (95% CI 0.4 to 7.8; $P=0.03$). Intensive care use was 343 of 1,255 (27.3%) versus 356 of 1,257 (28.3%), adjusted difference -1.0 percentage point (95% CI -4.4 to 2.5). Thirty-day mortality was 70 of 1,255 (5.6%) versus 64 of 1,257 (5.1%), adjusted difference 0.5

percentage points (95% CI –1.3 to 2.2). There were no statistically significant differences in emergency department use, early readmission or palliative care consultation. Mean time alive outside hospital was 21.6 versus 22.0 days, adjusted difference –0.4 days (95% CI –1.1 to 0.3).¹

Among minoritised patients, documented discussions occurred in 146 of 409 intervention patients (35.7%) versus 101 of 394 controls (25.6%), adjusted difference 10.2 percentage points (95% CI 4.0 to 16.5). Among non-Hispanic White patients, the corresponding figures were 276 of 815 (33.9%) versus 266 of 826 (32.2%), adjusted difference 1.6 percentage points (95% CI –3.0 to 6.2). The interaction P value was 0.03. No significant effect modification was identified for the other assessed characteristics. A sensitivity analysis expanding the definition of dementia through additional manual review was reported to give similar findings. Supplementary analyses examined separate racial and ethnic categories and the performance of the outcome-screening model.^{1,4}

Critique

PICSI-H establishes a modest improvement in documentation after a low-intensity communication prompt. The approximately four additional records per 100 patients are a useful process effect, but the endpoint does not establish that four additional patients received a high-quality conversation, that their families understood the options better, or that subsequent treatment matched their wishes more closely. The guide explicitly encouraged both discussion and documentation. Its two effects cannot be separated by an endpoint obtained solely from the record. This distinction is central to interpreting the trial, rather than a reason to dismiss documentation as unimportant.¹

Several design features support the primary comparison. Randomisation was individual, the analysis retained every allocated patient, and blinded human adjudication protected against consciously assigning more outcomes to the intervention group. Restricting eligibility to patients without an existing discussion focused the intervention on a recognisable gap in care. Consent waiver reduced the selection that can arise when only patients or families willing and able to complete research surveys participate. Nevertheless, clinician opt-outs and exclusions for certain clinical circumstances remained, and complete record-based follow-up is not equivalent to complete observation of conversations. An undocumented discussion could never satisfy the endpoint.¹

The measurement system deserves more scrutiny than a description of an automated outcome might suggest. Natural language processing screened text; people decided whether the selected passages met the endpoint definition. The reported standalone

model sensitivity of 99.2% and specificity of 33.2% describe its screening threshold, not the accuracy of the final human-adjudicated outcome. A permissive screen can appropriately generate many false-positive candidates if subsequent review removes them. The important remaining risk is missed qualifying material, because passages screened out do not receive that review. The combined method's 92.6% sensitivity was high but imperfect, and was established against manual review of documentation rather than direct observation of care.¹

The supplement provides a helpful check for differential model performance. In the 159-patient validation sample, patient-level discrimination did not differ significantly between allocation groups; nor was a significant difference identified by minoritised status among the 158 patients with that information. These post hoc analyses were small and deliberately oversampled people with dementia. Non-significant comparisons of areas under receiver operating characteristic curves do not prove equal sensitivity at the chosen threshold or establish identical error rates in every population. They reduce one concern without eliminating the possibility that language, documentation practices or differential recording influenced measured effects.⁴

The absolute primary effect is also imprecise relative to its modest size: the interval extends from less than one to almost eight additional documented discussions per 100 patients. The observed control proportion was lower than the average proportion assumed for planning, although the final sample exceeded the original target. The report explains the extension in terms of dementia recruitment and the endpoint of enrolment in terms of a forthcoming record-system change. It does not establish a chronology of access to comparative results that would justify labelling the extension either a response to emerging efficacy or definitively independent of it. The chapter therefore rests on the published analysis without inferring an undocumented stopping process.¹

Interpretation of the race and ethnicity finding requires both attention and restraint. The analysis used an interaction test, so the argument does not depend simply on one subgroup having a significant result and another not. The larger absolute gain among minoritised patients is clinically relevant and merits confirmation. However, several modifiers were examined without a reported multiplicity adjustment, 68 patients lacked the combined classification, and a single pooled category contains groups with different experiences. The supplement's separate estimates are often very imprecise, especially in the smallest categories. These data cannot identify which barriers the prompt addressed or establish an equal benefit for every racial or ethnic group. They support further evaluation of equitable communication, not a biological interpretation of race or a rule restricting the guide to selected patients.^{1,4}

The dementia analysis is similarly limited by precision. Only 140 patients per group had dementia identified at allocation, and their estimated difference was –1.4 percentage points with an interval from –12.8 to 9.9. This does not demonstrate lack of benefit in dementia. The guide was sent to clinicians and did not include a separately tested intervention to prepare family decision-makers. Whether a different approach is needed when capacity, memory or surrogate involvement changes the conversation remains unresolved.¹

Delivery fidelity is an important practical uncertainty. Sending a guide is not the same as its being read, judged relevant, used at the bedside or shared during handover. Without those measurements, the trial cannot distinguish limited engagement from limited efficacy of the conversation prompts themselves. Individual randomisation also permits the same clinician to care for patients in both groups. Learning from an intervention patient's guide could influence usual care and reduce the observed contrast, but the extent of any such spillover is unknown. Neither explanation licenses replacing the observed effect with an assumed larger effect under ideal use.¹

The absence of differences in intensive care use, hospital-free time or mortality does not prove equivalence or establish safety. These outcomes were secondary, and their intervals include effects in both directions. More fundamentally, reduced intensive care use is not automatically a desirable consequence of a goals discussion: treatment consistent with a patient's priorities may appropriately include intensive care. Patient-reported communication quality, distress and concordance between goals and care were not measured in this trial. It therefore provides no direct assessment of potential emotional harms or of the patient benefit that would justify regarding documentation as a successful intervention in its own right.¹

Generalisability concerns are specific rather than merely geographical. Three different hospitals broaden the clinical range, but all used one health system's infrastructure and documentation practices during a pandemic year. Events and conversations elsewhere could be missed. The trial selected older patients with relevant diagnoses or advanced age, not all adults at risk of deterioration, and many had relatively low short-term mortality. For an intensive care service, the applicable lesson is that an early hospital-team prompt can improve a recorded process. The trial does not establish how well the same approach works during emergency intubation, prolonged critical illness or a difficult treatment-withdrawal discussion.¹

Body of Evidence

The earlier outpatient Jumpstart trial randomised 132 clinicians and included 537 patients with serious illness. Its original abstract reports patient-identified goals-of-care discussions in 74% versus 31%, with documented discussions in 62% versus 17% and better reported communication quality. Overall goal-concordant care did not differ significantly. Both patients and clinicians received preparation, so this was a more involved intervention than PICS-H. Its large communication effect supplies a rationale for priming, but is not a directly transferable estimate for a clinician-only hospital email.²

The subsequent hospital pilot randomised 150 patients, with guides for patients or surrogates as well as clinicians. Documented discussions between randomisation and discharge occurred in 16 of 75 intervention patients and six of 75 controls, an absolute difference of 13 percentage points (95% CI 2 to 24). Patient or surrogate reports of discussions and communication quality did not improve significantly. This discrepancy already showed why a documentary endpoint and a reported experience cannot be treated as interchangeable. The pilot also used a different observation window and a consented, substantially younger cohort.³

An independent oncology trial tested a different way of prompting clinicians. Manz and colleagues used a stepped-wedge cluster design across nine clinics, combining mortality-risk information with behavioural reminders and feedback. The original abstract reports documented serious-illness conversations at 4.6% versus 1.3% of encounters, an adjusted increase of 3.3 percentage points. This provides convergent evidence that changing how clinicians are prompted can change recording behaviour. Its encounter-based denominator, cancer population and multicomponent intervention distinguish it from a hospital patient-level trial. A published correction affected high-risk subgroup calculations; those estimates are not used here.⁵

The companion natural language processing study addresses feasibility of measurement rather than providing independent confirmation of PICS-H's clinical effect. It used the same 2,512-patient trial cohort, with an internal validation sample. Human review of 7,494 screened passages required 34.3 abstractor hours. This illustrates a substantial reduction in the amount of text requiring review, but is not an estimate of the total cost of developing, validating or operating the intervention and measurement system. The method still depends on an appropriate documentation standard and human adjudication.⁶

Together, these studies describe two different questions. One is whether clinicians and patients can be helped to initiate and record discussions; the randomised evidence is encouraging. The other is whether a particular scalable implementation improves

outcomes that patients recognise and value. PICS-H advances the first question by testing a simpler delivery model in a larger and less selectively consented hospital population. It leaves the second largely unanswered. Neither a smaller effect than an earlier trial nor a larger usual-care documentation rate isolates the consequences of removing patient preparation, because populations, outcome windows and measurement methods changed simultaneously.^{1,2,3}

For practice at the end of 2023, the guide is best understood as an aid to organising communication, with a measurable but limited documentary benefit. A service adopting a similar approach would need to determine whether clinicians engage with it, whether patients and families recognise the resulting discussions, and whether the record conveys priorities usefully to subsequent teams. Evaluation should preserve the distinction between an appropriate offer of conversation, its quality, its documentation and the care eventually delivered. Maximising a documentation percentage alone could reward brief formulaic entries without resolving the decision that matters to the patient.

PICS-H consequently supports testing simple prompts within routine hospital workflows, particularly where relevant information exists but conversations are easily deferred. It does not establish that more elaborate preparation is unnecessary, that usual communication can be replaced by a form, or that the larger effect in the pooled minoritised population will reproduce elsewhere. The next substantive step is to connect improvements in recorded communication to patient and family experience while retaining the broad reach that made this pragmatic approach attractive.

Does a clinician-facing Jumpstart Guide improve goals-of-care communication for older hospital patients?

It modestly increases documented discussions. Better conversation quality, care aligned with patients' goals and improved clinical outcomes remain unproven.

References

1. Curtis JR, Lee RY, Brumback LC, Kross EK, Downey L, Torrence J, et al. Intervention to Promote Communication About Goals of Care for Hospitalized Patients With Serious Illness: A Randomized Clinical Trial. *JAMA*. 2023;329(23):2028–2037.
2. Curtis JR, Downey L, Back AL, Nielsen EL, Paul S, Lahdya AZ, et al. Effect of a Patient and Clinician Communication-Priming Intervention on Patient-Reported Goals-of-Care Discussions Between Patients With Serious Illness and Clinicians: A Randomized Clinical Trial. *JAMA Intern Med*. 2018;178(7):930–940.

3. Lee RY, Kross EK, Downey L, Paul SR, Heywood J, Nielsen EL, et al. Efficacy of a Communication-Priming Intervention on Documented Goals-of-Care Discussions in Hospitalized Patients With Serious Illness: A Randomized Clinical Trial. *JAMA Netw Open*. 2022;5(4):e225088.
4. Curtis JR, Lee RY, Brumback LC, Kross EK, Downey L, Torrence J, et al. Supplemental online content to: Intervention to Promote Communication About Goals of Care for Hospitalized Patients With Serious Illness: A Randomized Clinical Trial. *JAMA*. 2023.
5. Manz CR, Parikh RB, Small DS, Evans CN, Chivers C, Regli SH, et al. Effect of Integrating Machine Learning Mortality Estimates With Behavioral Nudges to Clinicians on Serious Illness Conversations Among Patients With Cancer: A Stepped-Wedge Cluster Randomized Clinical Trial. *JAMA Oncol*. 2020;6(12):e204759.
6. Lee RY, Kross EK, Torrence J, Li KS, Sibley J, Cohen T, et al. Assessment of Natural Language Processing of Electronic Health Records to Measure Goals-of-Care Discussions as a Clinical Trial Outcome. *JAMA Netw Open*. 2023;6(3):e231204.

Summarising, critiquing and putting into context the best critical care trials of 2023



Rob Mac Sweeney