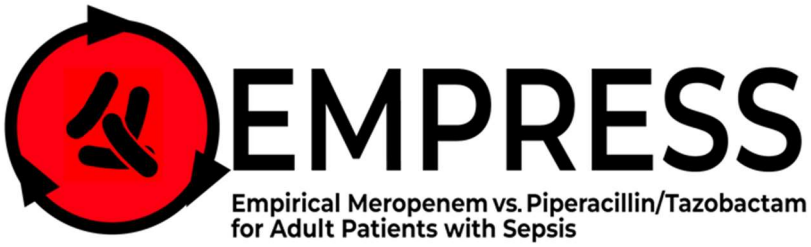


Empirical Meropenem versus Piperacillin/Tazobactam for Adult Patients with Sepsis (EMPRESS) trial

Management committee

Morten Hylander Møller, sponsor
Anders Perner
Anders Granholm
Marie Warrer Munch
Nick Meier
Theis Lange
Aksel Karl Georg Jensen
Benjamin Skov Kaas-Hansen
Maj-Brit Nørregaard Kjær
Fredrik Sjövall
Marie Helleberg
Frederik Boëtius Hertz
Bodil Steen Rasmussen
Lars Wiuff Andersen
Hans-Christian Thorsen-Meyer
Jakob Steen Andersen
Trine Lynge Albertsen

Protocol version and date

Version 1.10, 2024.03.22

Websites

www.cric.nu/empress
www.incept.dk/empress

Applicable protocol registration numbers

Clinical Trials Information System (CTIS)
EUCT number: 2023-509703-33-00
ClinicalTrials.gov identifier: NCT06184659
Universal Trial Number: U1111-1301-6379

1 Abstract

Background

Piperacillin/tazobactam and meropenem are commonly used as empirical agents for patients with sepsis or septic shock. In a recent systematic review comparing empirical and/or definitive treatment with piperacillin/tazobactam versus meropenem for patients with severe bacterial infections, including sepsis and septic shock, it was shown that piperacillin/tazobactam may be associated with less favourable outcomes based on low or very low certainty of evidence. At present, it is unclear if piperacillin/tazobactam and meropenem are equally effective and safe for adults with sepsis.

Objectives

To assess the effects of empirical meropenem versus piperacillin/tazobactam on mortality and other patient-important outcomes in critically ill adults with sepsis.

Design

Investigator-initiated, international, parallel-group, randomised, open-label, adaptive clinical trial with an integrated feasibility phase. The EMPRESS trial will employ adaptive stopping rules to increase the chance that the trial will be conclusive and response-adaptive randomisation to increase each participant's chance of being randomised to the superior intervention arm.

Inclusion and exclusion criteria

We will screen all adult patients who are critically ill with sepsis and who have indication for empirical treatment with meropenem or piperacillin/tazobactam. We will exclude patients with preceding intravenous treatment with meropenem or piperacillin/tazobactam for 24 hours or more; known pregnancy; known hypersensitivity or allergy to beta-lactam antibiotics; suspected or documented central nervous system infection; known infection or colonialisation with microorganism with acquired resistance to meropenem or piperacillin/tazobactam; current or planned use of valproate; co-enrolment in other interventional trial where protocols collide; previous randomisation in EMPRESS; informed consent following inclusion expected to be unobtainable; and patients who are under coercive measures.

Experimental intervention

IV meropenem 1 g three times daily for up to 30 days.

Control intervention

IV piperacillin/tazobactam 4/0.5 g four times daily for up to 30 days.

Outcomes

The primary outcome is all-cause mortality at 30 days after randomisation. The secondary outcomes are the occurrence at least one serious adverse reaction (i.e., anaphylactic reaction to IV piperacillin/tazobactam or meropenem, invasive fungal infection, pseudomembranous colitis, or toxic epidermal necrolysis) within 30 days of randomisation; the occurrence of new isolation precautions due to resistant bacteria within 30 days of randomisation; days alive without life support (i.e., invasive mechanical ventilation, circulatory support, or renal replacement therapy) from randomisation to day 30 and 90; days alive and out of hospital from randomisation to day 30 and 90; all-cause mortality at day 90 and 180; and health-related quality of life at day 180 using EQ-5D-5L index values and EQ VAS. The feasibility outcomes are time to completion of feasibility phase (i.e., 200 participants randomised), recruitment proportion, proportion of participants without consent to the use of data, protocol adherence, and retention proportion.

Statistics and stopping rules

The trial will be analysed using Bayesian statistical methods with the primary analyses conducted in the intention-to-treat population. Outcomes will be analysed using logistic and linear regression models adjusted for relevant baseline characteristics and neutral and weakly informative to somewhat sceptical priors. Results will be presented as adjusted sample average treatment effects using both absolute (risk and mean differences) and relative (risk ratios and ratios of means) differences with 95% credible intervals and probabilities of benefit/harm. Adaptive analyses will start after follow-up and data collection concludes for 400 participants and every subsequent 300 participants up to a maximum of 14,000 participants. Adaptations will be based on data for the primary outcome. EMPRESS will use constant, symmetrical stopping rules for inferiority/superiority calibrated to keep the type 1 error rate at 5%. Further, the trial will be stopped for practical equivalence if there is >90% probability that the absolute risk difference between arms is less than 2.5%-points. Restricted response-adaptive randomisation will be used to ensure minimum allocation probabilities of 40% to both groups.

Missing data will be imputed, and relevant secondary analyses, sensitivity analyses, and analyses of heterogeneity in treatment effects according to pre-defined baseline characteristics will be undertaken once the trial has been stopped.

Trial design performance metrics

Performance characteristics were evaluated assuming a 25% event probability for the primary outcome in the piperacillin/tazobactam arm and scenarios with no, small, and large differences corresponding to event probabilities of 25%, 22.5%, and 20% in the meropenem arm, respectively. The expected (mean) sample sizes under these scenarios are 5189, 5859, and 2570, respectively. The probabilities of conclusiveness (i.e., superiority or equivalence) are >99% in all scenarios, and the probabilities of superiority (power) are 72% and >99% in the small and large difference scenarios, respectively.

Estimated timeline

- Medio 2024: authority approvals and first participant randomised
- Medio 2025: feasibility phase analysis concluded
- Ultimo 2028: expected inclusion of the last participant if trial continues to the expected sample size in the small-difference scenario (i.e., the largest expected sample size under the three different scenarios assessed) (section 19.9)
- Ultimo 2032: expected inclusion of the last participant if the trial continues to the maximum sample size (n=14,000) (section 19.9)
- Approximately 3 months after inclusion of the last participant: primary report on 30-day outcomes submitted
- Approximately 6 months after inclusion of the last participant: report on 90-day outcomes submitted
- Approximately 9 months after inclusion of the last participant: report on 180-day outcomes submitted

2 Administrative information

2.1 Sponsor and coordinating centre

International central coordinating centre

Department of Intensive Care

Rigshospitalet

Blegdamsvej 9

2100 Copenhagen Ø

+45 3545 7167

contact@cric.nu

International sponsor

Morten Hylander Møller, senior staff specialist in intensive care medicine and professor

Department of Intensive Care

Rigshospitalet

Blegdamsvej 9

2100 Copenhagen Ø

+45 3545 8685

morten.hylander.moeller@regionh.dk

National sponsors and coordinating centres

A list of national sponsors and coordinating centres will be available with the registration in the Clinical Trials Information System (CTIS) prior to trial initiation and continuously updated during the trial.

2.2 Local investigators and clinical trial sites

A list of local investigators and clinical trial sites will be available with the registration in the Clinical Trials Information System (CTIS) prior to trial initiation and continuously updated during the trial.

2.3 Methodological sites

Coordinating centre

Department of Intensive Care

Rigshospitalet

Blegdamsvej 9

2100 Copenhagen Ø

+45 3545 0606

empress@cric.nu

Methodological and statistical centres

Department of Intensive Care

Rigshospitalet

Blegdamsvej 9

2100 Copenhagen Ø

Section of Biostatistics

University of Copenhagen

Øster Farimagsgade 5

1014 Copenhagen K

Data management centre

sundhed.dk

Dampfærgevej 22

2100 København Ø

2.4 Independent data monitoring and safety committee (IDMSC)

The charter for the IDMSC is included in appendix 1, section 19.1. The members of the IDMSC are listed below.

IDMSC trialist (chair)

Kathy Rowan, The Intensive Care National Audit and Resource Centre (ICNARC)

London, United Kingdom

IDMSC clinician

Lennie Derde, Department of Intensive Care Medicine, UMC Utrecht

Utrecht, The Netherlands

IDMSC statistician

Erin Evelyn Gabriel, Section of Biostatistics, Department of Public Health University of Copenhagen, Copenhagen, Denmark

3 List of abbreviations

ACT-EU: *Accelerating Clinical Trials in the EU*

ADRENAL: *Adjunctive Glucocorticoid Therapy in Patients with Septic Shock trial*

AE: *adverse event*

AR: *adverse reaction*

ATC: *anatomical therapeutic chemical classification system*

CI: *confidence interval*

CLASSIC: *Conservative vs. Liberal Approach to Fluid Therapy of Septic Shock in Intensive Care trial*

CONSORT: *Consolidated Standards Of Reporting Trials*

CONSORT-ACE: *Adaptive Designs Consolidated Standards Of Reporting Trials Statement Extension*

COVID STEROID 2: *Higher vs. Lower Doses of Dexamethasone in Patients with COVID-19 and Severe Hypoxia trial*

COVID STEROID: *Low-Dose Hydrocortisone in Patients with COVID-19 and Severe Hypoxia trial*

CPAP: *continuous positive airway pressure*

CrI: *credible interval*

CRIC: *Collaboration for Research in Intensive Care*

CRP: *C-reactive protein*

CT: *computed tomography*

CTIS: *Clinical Trials Information System*

DeIC: *Danish e-infrastructure Consortium*

DIANA: *Antimicrobial De-escalation in the Critically Ill Patient and Assessment of Clinical Cure study*

ECG: *electrocardiogram*

eCRF: *electronic case report form*

EMA: *European Medicines Agency*

EMPRESS: *Empirical Meropenem vs. Piperacillin/Tazobactam for Adult Patients with Sepsis trial*

ESBL: *extended spectrum beta-lactamase*

EU: *European Union*

EUCAST: *European Committee of Antimicrobial Susceptibility Testing*

FDA: *the United States Food and Drug Administration*

FiO_2: *Fraction of Inspired Oxygen*

GCP: *Good Clinical Practice*

GRADE: *the Grading of Recommendations Assessment, Development and Evaluation approach*

hCG: *human chorionic gonadotropin*

HPC: *high-performance computing*

HR: *hazard ratio*

HRQoL: *health-related quality of life*

HTE: *heterogenous treatment effects*

ICH-GCP: *International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use guidelines on Good Clinical Practice*

ICMJE: *International Committee of Medical Journal Editors*

ICU: *intensive care unit*

IDMSC: *independent data monitoring and safety committee*

IDP: *ideal design percentage*

IL: *interleukin*

INCEPT: *Intensive Care Platform Trial*

IQR: *interquartile range*

ITT: *intention-to-treat*

IV: *intravenous*

kg: *kilogram*

L: *litre*

m: *meter*

MAR: *missing at random*

MCAR: *missing completely at random*

MD: *mean difference*

MERINO: *Meropenem versus Piperacillin/tazobactam for Definitive Treatment of Bloodstream Infections due to Ceftriaxone Non-Susceptible* Escherichia Coli *and* Klebsiella Spp. *trial*

mg: *milligram*

MI: *multiple imputation*

mL: *millilitre*

mmHg: *millimetres of mercury*

mmol: *millimole*

MRSA: *methicillin-resistant* Staphylococcus aureus

MSSA: *methicillin-sensitive* Staphylococcus aureus

NMAR: *not missing at random*

OR: *odds ratio*

P25: *25th percentile*

P75: *75*th *percentile*

PaO_2: *partial pressure of arterial oxygen*

PCR: *polymerase chain reaction*

Pr: *probability*

RCT: *randomised clinical trial*

RMSE: *root mean squared error*

ROBUST: *Reporting Of Bayes Used in clinical STudies*

RoM: *ratio of means*

RR: *relative risk*

SAE: *serious adverse event*

SaO_2: *arterial oxygen saturation*

SAR: *serious adverse reaction*

SD: *standard deviation*

SmPC: *summary of product characteristics*

SMS-ICU: *Simplified Mortality Score for the Intensive Care Unit*

SOFA: *Sequential Organ Failure Assessment score*

SPIRIT: *Standard Protocol Items: Recommendations for Interventional Trials*

SpO_2: *peripheral oxygen saturation*

Spp.: *species*

SUSAR: *suspected unexpected serious adverse reaction*

VRE: *Vancomycin-resistant* Enterococci

µmol: *micromole*

4 Introduction and background

4.1 Sepsis

Sepsis is characterised by life-threatening organ dysfunction caused by a dysregulated host response to infection (1). Sepsis may progress to septic shock with end-organ dysfunction resulting in elevated lactate levels and persistent hypoperfusion requiring fluids and vasopressor treatment (1). Sepsis is associated with high mortality, and the progression to septic shock results in worsening of outcomes, including even higher risk of mortality (2). Of those who survive, many suffer from long-term sequela and reduced quality of life (3).

The global burden of sepsis is high with approximately 50 million cases of sepsis and 11 million sepsis-related deaths every year (4). This makes sepsis one of the leading causes of deaths globally (4, 5). The resources used to treat sepsis is also high. It is estimated that 30-40% of patients admitted to the intensive care unit (ICU) in high income countries have sepsis (2, 3). Consequently, there is an urgent need to optimise the treatment to improve the outcomes for this vulnerable group of patients.

4.2 Broad-spectrum antimicrobial treatment for sepsis and septic shock

Early identification and initiation of treatment improve outcomes for patients with sepsis and septic shock (2). Updated recommendations for the care of patients with sepsis and septic shock are described in the Surviving Sepsis Campaign 2021 guidelines (2). In this, it is recommended to administer antimicrobials immediately for patients with possible septic shock or a high likelihood for sepsis (strong recommendation based on low [for septic shock] and very low [for sepsis without shock] certainty evidence) (2). For patients with possible sepsis without shock, these guidelines suggest that antimicrobials be administered within 3 hours if the concern of infection persists (conditional recommendation based on very low certainty evidence) (2). It is suggested to use an antimicrobial agent with gram-positive as well as gram-negative coverage as empirical treatment (conditional recommendation based on very low certainty evidence) (2).

4.3 Type, dose, and formulation of antimicrobial agents for sepsis and septic shock

Type of antimicrobial agent

The choice of empirical antimicrobial agents depends on various factors, including primary source of infection, patient comorbidities, and local resistance patterns, but should generally target the most likely and most harmful potential pathogens (3).

Piperacillin/tazobactam (a β-lactam/β-lactamase inhibitor) and meropenem (a carbapenem) are commonly used as empirical agents for patients with sepsis or septic shock (6, 7). In an international retrospective cohort of 3800 patients with septic shock included in the *Adjunctive Glucocorticoid Therapy in Patients with Septic Shock (ADRENAL)* trial (6) and 1495 critically ill patients receiving empirical antibiotics in the ICU included in the *Antimicrobial De-escalation in the Critically Ill Patient and Assessment of Clinical Cure (DIANA)* study (7), 49% and 29% received initial treatment with piperacillin/tazobactam, respectively, and 17% and 26% received initial treatment with carbapenems (e.g., meropenem), respectively (8). In a local cohort of 286 patients admitted to the Department of Intensive Care at Rigshospitalet for 12 consecutive weeks during 2022-2023, meropenem and piperacillin/tazobactam were used in 36% and 38% of patients during their ICU admission, respectively (unpublished data from our group).

Piperacillin/tazobactam are often preferred over meropenem as carbapenems have been associated with an increased risk of multidrug-resistant pathogens and superinfections (9, 10). In the non-inferiority MERINO trial, piperacillin/tazobactam was compared with meropenem for the definitive treatment of patients with bloodstream infection with ceftriaxone-non-susceptible *Escherichia coli* or *Klebsiella* species (*spp.*) (11, 12). The 30-mortality was markedly higher in the group of patients receiving piperacillin/tazobactam and the trial concluded that piperacillin/tazobactam was not non-inferior to meropenem (11, 12).

Previous systematic reviews comparing β-lactam/β-lactamase inhibitors and carbapenems for Extended spectrum β-lactamase (ESBL)-producing *Enterobacterales* bacteremia (13-16), AmpC-producing *Enterobacterales* (17, 18), and other severe infections (19-22) have shown conflicting results. The most recent systematic review of randomised clinical trials comparing empirical and/or definitive treatment with piperacillin/tazobactam versus carbapenems for adults with severe bacterial infections (i.e., any infection requiring hospitalisation) found that piperacillin/tazobactam may be associated with higher all-cause short-term mortality based on low certainty of evidence, although a potential small mortality benefit with piperacillin/tazobactam could not be ruled out (**Table 1**) (22). This was not explained by a difference in the

occurrence of adverse events, which were similar with both drugs (moderate certainty evidence) (22). The effects on secondary infections and selection of fungi or resistant bacteria were uncertain (very low certainty of evidence) (22). In the subgroup of patients receiving empirical piperacillin/tazobactam versus carbapenems, the evidence suggested that piperacillin/tazobactam results in little to no difference in all-cause short-term mortality, but the confidence interval was wide (i.e., included both substantial potential benefit and harm from piperacillin/tazobactam) (22). Thus, at present, it is unclear if piperacillin/tazobactam and meropenem are equally effective and safe for patients with sepsis (22).

Table 1. Estimates on the effects of carbapenems versus piperacillin/tazobactam in adult patients with severe infection and bloodstream infections.

	Severe infections requiring hospitalisation
Evidence base	Systematic review with meta-analysis of 31 RCTs involving 8790 patients (all with overall high risk of bias) (22)
Types of infection	Intra-abdominal infections: 7 trials Pneumonia: 6 trials Febrile neutropenia: 5 trials Skin- and skin-structure infections: 4 trials Urogenital infections: 4 trials Bloodstream infections: 2 trials Mixed bacterial infections: 2 trials Combined population of intra-abdominal infections and pneumonia: 1 trial
Publication year of included trials	1993 to 2001
All-cause short-term mortality (≤ 90 days)	Relative risk: 0.86 (95% CI: 0.70 to 1.06) Absolute effect: 10 fewer (95% CI: 20 fewer to 4 more) per 1000 patients Low certainty of evidence <u>Subgroup analysis of RCTs assessing empirical treatment only (18 trials, 5778 patients):</u> Relative risk: 1.01 (95% CI: 0.79 to 1.27) Absolute effect: 0 (95% CI: 10 fewer to 12 more) per 1000 patients
Adverse events	Relative risk: 1.00 (95% CI: 0.96 to 1.04) Absolute effect: 0 (95% CI: 16 fewer to 16 more) per 1000 patients Moderate certainty of evidence
Secondary infections	Relative risk: 0.88 (95% CI: 0.61 to 1.30) Absolute effect: 6 fewer (95% CI: 19 fewer to 15 more) per 1000 patients Very low certainty of evidence
Selection of fungi or resistant bacteria	Relative risk: 0.63 (95% CI: 0.38 to 1.03) Absolute effect: 17 fewer (95% CI: 28 fewer to 1 more) per 1000 patients Very low certainty of evidence

Dose and formulation of piperacillin/tazobactam and meropenem

In the latest guidance from the European Committee of Antimicrobial Susceptibility Testing (EUCAST), the recommended standard dose of piperacillin/tazobactam is 4/0.5 g administered 4 times daily as a 30-minute intravenous (IV) infusion or 3 times daily by extended 4-hour infusions and the recommended high dose is 4/0.5 g administered 4 times daily by extended 3-hour infusion (23). For meropenem, the standard dose is 1 g administered 3 times daily as a 30-minute intravenous (IV) infusion, and the high dose is 2 g administered 3 times daily by extended 3-hour infusion (23). The same dose is recommended in the guidelines for the treatment of sepsis by the Danish Society of Infectious Diseases (24).

Type and dose of empirical antimicrobial agents in the EMPRESS trial

In the *Empirical Meropenem vs. Piperacillin/Tazobactam for Adult Patients with Sepsis (EMPRESS) trial*, participants in the experimental intervention arm will receive 1 g of IV meropenem three times daily administered according to usual clinical practice (i.e., intermittent-, prolonged-, or continuous infusion) until discharge from the participating site, death, termination of empirical antibiotic therapy (including initiation of definitive treatment), or de-escalation to another empirical antibiotic with more narrow spectrum (not including the two trial interventions), i.e., in situations where no bacteria are found in microbiological cultures. The dose of meropenem will be adjusted according to pre-defined criteria based on certain patient and infection characteristics (section 8.4).

Participants in the control arm will receive 4/0.5 g of intravenous piperacillin/tazobactam four times daily administered according to usual clinical practice (i.e., intermittent-, prolonged-, or continuous infusion) until discharge from the participating site, death, termination of empirical antibiotic therapy (including initiation of definitive treatment) , or de-escalation to another empirical antibiotic with more narrow spectrum (not including the two trial interventions), i.e., in situations where no bacteria are found in microbiological cultures. The dose of piperacillin/tazobactam will likewise be adjusted according to pre-defined criteria (section 8.5).

The trial medication will be continued for maximum 30 days. If there is clinical indication for empirical broad-spectrum antibiotic treatment for subsequent infections after initial de-escalation/discontinuation of the intervention but within 30 days of randomisation, these subsequent infections will be treated empirically according to the allocation. Hereafter, the choice of antibiotics is up to the treating clinician, although it is recommended that the allocated intervention be continued if relevant. We will allow antibiotic combination therapy in both arms.

4.4 Ethical justification and trial rationale

Patients with sepsis and septic shock are at high risk of death (2). One key recommendation for the treatment of these patients is early administration of empirical broad-spectrum antimicrobial agents (2).

The MERINO trial questioned the efficacy and safety of definitive treatment with piperacillin/tazobactam as compared with meropenem in patients with bloodstream infection with ceftriaxone-non-susceptible *E. coli* or *Klebsiella* spp. (11, 12). In the most recent systematic review comparing empirical and/or definitive treatment with piperacillin/tazobactam versus meropenem for patients with severe bacterial infections, it was concluded that piperacillin/tazobactam may be associated with less favourable outcomes, but the certainty of evidence was very low or low for most outcomes (**Table 1**) (22). The subgroup analysis of RCTs assessing empirical treatment only suggested that piperacillin/tazobactam results in little to no difference in all-cause short-term mortality, but the confidence interval included both benefit and harm of piperacillin/tazobactam. Of the 31 eligible trials, none were adjudicated overall low risk of bias (22). Moreover, there were no trials exclusively done in critically ill patients (22), and thus the mortality rates with both treatments were substantially lower than in the critical care setting. These findings highlight the need for a high-quality trial done in critically ill patients with sepsis.

The present trial will be conducted according to high methodological standards, including an integrated feasibility phase and ongoing assessment of the known serious adverse reactions to meropenem and piperacillin/tazobactam. Any serious adverse reactions for single participants and the group of participants receiving meropenem or piperacillin/tazobactam will be assessed and handled. Both interventions will be administered in addition to usual clinical care. Importantly, the adaptive design will increase the likelihood that the trial results will be conclusive and thus inform clinical guidelines and practice (25). A detailed rationale for conducting the trial as an adaptive, Bayesian trial is provided in section 12.1.

The trial will be conducted according to the applicable laws in the participating countries for clinical trials conducted in emergency situations (including the European [EU] and Danish legislation (26, 27)). The patients eligible for enrolment in the EMPRESS trial cannot consent due to critical illness from severe infection and organ dysfunction (e.g., hypotension, hypoxia, delirium). Sepsis and septic shock are medical emergencies that requires immediate interventions, including early administration of broad-spectrum antibiotics. Therefore, we cannot delay enrolment and need to use the consent procedures for emergency research. Informed consent will be obtained according to the applicable laws in the participating countries,

with enrolment and informed consent procedures handled by trained trial staff with thorough insight into the protocol.

4.5 Trial conduct

The EMPRESS trial will comply with the published trial protocol, the Helsinki Declaration in its latest version (28), the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use guidelines on Good Clinical Practice (ICH-GCP) guidelines (29), General Data Protection Regulation, and national laws (including Databeskyttelsesloven in Denmark). The management committee of the trial will oversee the conduct. We have written the protocol in accordance with the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) 2013 Statement (30), and the relevant items from the Adaptive designs Consolidated Standards Of Reporting Trials (CONSORT) statement Extension (CONSORT-ACE), which covers the reporting of adaptive designs and is used here as no specific checklist for planning adaptive designs exist (31). Both completed checklists are included in appendix 2, section 19.2. We will register the trial in the Clinical Trials Information System (CTIS) and ClinicalTrials.gov registers before enrolling the first participant. No substantial deviation from the protocol will be implemented without prior review and approval of the regulatory authorities except in case of urgent safety measures where it may be necessary to eliminate an immediate hazard to the trial participants. In such case, the deviation will be reported to the authorities within 7 days (15 days if the deviation leads to pausing of the trial) (32).

Enrolment will start after the approval by the competent authorities. We will publish the approved protocol online on the trial website and submit a manuscript with main points of the protocol, including description of design, rationale, and the detailed statistical analysis plan to a peer-reviewed medical journal.

5 Trial objectives

The objective of the *Empirical Meropenem vs. Piperacillin/Tazobactam for Adult Patients with Sepsis (EMPRESS) trial* is to assess the effects of empirical meropenem versus piperacillin/tazobactam on mortality and other patient-important outcomes in critically ill adults with sepsis. We hypothesise that meropenem will lower mortality compared to piperacillin/tazobactam without any substantial effects on adverse outcomes.

6 Trial design

6.1 Overall design and sample size

The EMPRESS trial is an investigator-initiated, international, parallel-group, randomised, open-label, adaptive clinical trial with an integrated feasibility phase. The EMPRESS trial will employ adaptive stopping rules (for superiority/inferiority/practical equivalence, and with a maximum pre-specified sample size) to increase the chance that the trial will be conclusive and response-adaptive randomisation to increase each participant's chance of being randomised to the superior intervention arm (25). The maximum sample size will be 14,000 participants as this is necessary to ensure >99% probabilities of the trial ultimately being conclusive, i.e., fulfilling either the criterion for superiority of one intervention or the criterion for practical equivalence of both interventions across a range of plausible clinical scenarios (detailed in section 12.6 and appendix 9, section 19.9). The expected (i.e., most realistic) sample size across the evaluated clinical scenarios ranges from 2,570 to 5,859 participants (section 12.6). Based on a conservative estimate of an expected inclusion rate of 5 participants per day (informed by previous trials by our group (33-37)) after the feasibility phase consisting of 200 participants (section 6.3) , the trial is expected to conclude enrolment within reasonable time, i.e., after approximately 4.5 years if the final sample size matches the upper range of the expected sample sizes across clinical scenarios and approximately 8.5 years if the trial continues to the maximum sample size (section 16). These scenarios are expected to be feasible by the trial group.

EMPRESS is a stand-alone clinical trial that is partially based on the methodology and core protocol for the upcoming adaptive platform trial INCEPT (the Intensive Care Platform Trial, www.incept.dk) sharing parts of the methodological framework and infrastructure being developed for INCEPT and conducted within the Collaboration for Research in Intensive Care (CRIC, www.cric.nu) network.

6.2 Randomisation

Critically ill adult patients with sepsis admitted to participating sites and fulfilling all inclusion criteria and no exclusion criteria will be randomised using a centralised and web-based randomisation system. We will use response-adaptive randomisation to adapt allocation ratios at regularly conducted adaptive analyses throughout the trial with equal (1:1) allocation ratios prior to the first adaptive analysis (i.e., first analysis after the feasibility phase). As response-adaptive randomisation increases the probability of participants being allocated to the treatment arm with the highest probability of being the best, it may be preferable for patients *internal* to the trial (i.e., those included in EMPRESS) (25). However, depending on the exact trial

design, response-adaptive randomisation may also increase the total sample size required (due to, e.g., fewer total events in the trial, large deviations from equal allocation, or adaptation to random fluctuations) leading to longer time until trial results can benefit patients *external* to the trial (25). While unequal randomisation may be less effective than equal randomisation in trials with continuous primary outcomes, unequal randomisation may make trials more effective (i.e., increase power) when assessing binary outcomes due to more similar total event counts in both arms. EMPRESS will use *restricted response-adaptive randomisation* to avoid too dissimilar allocation ratios in the two groups with the exact restrictions evaluated by statistical simulation to adequately balance benefits to *internal* and *external* patients as described in sections 12.5 and 12.6 (25). The use of restricted randomisation also mitigates the risk of over-aggressive adaptations to random fluctuations (30), and restricted response-adaptive randomisation is presently used in other large-scale two-arm critical care trials (38). Finally, restricted response-adaptive randomisation also makes it somewhat more difficult for clinicians and investigators to guess which treatment is most likely to be superior prior to a stopping rule being reached, which may otherwise affect clinical equipoise and willingness to randomise patients.

Randomisation will initially be stratified by trial site using computer-generated allocation sequence lists with 1:1 ratios and randomly varying block sizes (39); these allocation sequences lists will be prepared by a person not involved in patient inclusion or outcome data collection, and will be concealed with allocations revealed immediately after randomisation. Following the first adaptive analysis we will change to response-adaptive randomisation, using simple (unstratified) randomisation (39) according to the response-adaptive randomisation allocation ratios active at the time of each randomisation; here, the allocation will be generated immediately and revealed by a computer system at the time of inclusion. This is done as stratified blocked randomisation is infeasible with frequently updated allocation probabilities where the allocation ratios between arms will often require very large blocks to obtain the intended ratio. Each trial participant will be allocated a unique screening number.

6.3 Feasibility phase

The EMPRESS trial will include an initial, integrated feasibility phase. Feasibility will be assessed after data collection for all feasibility outcomes (section 9.2) for the first 200 participants. While clinical outcomes will be collected, only feasibility outcomes will be analysed during the feasibility phase according to the criteria and considerations outlined in section 9.2. Of note, feasibility outcomes will be reported for the full population and for the two groups separately, with overall feasibility primarily determined based on the full population. Inclusion and delivery of the trial interventions will continue according to the approved

protocol until a final decision from the feasibility phase has been made and implemented.
The sample size for the integrated feasibility phase has been determined based on practical considerations (40), sample sizes in other pilot/feasibility trials (41), and the expected precision for the estimates of the feasibility outcomes (section 9.2) (42).

The trial will either continue unaltered (if all feasibility outcomes have been fulfilled); with modifications to the protocol as deemed necessary (if some feasibility outcomes have not been fulfilled); or stopped if deemed infeasible (at the discretion of the management committee after consulting the independent data monitoring and safety committee [IDMSC]). If changes to the protocol are deemed necessary, inclusion in the trial will be paused (if required) until the competent authorities have approved the applied changes. The trial will not pause inclusion while feasibility outcomes are being collected or evaluated.

No analyses of the clinical outcomes will be conducted before the feasibility phase has concluded. If the trial is stopped due to infeasibility, all included participants at that time-point will continue receiving the allocated trial interventions and follow-up, and analyses of clinical outcomes by treatment allocation will be conducted after follow-up has been completed for all included participants. If the trial is modified after the feasibility phase, a subsequent assessment of feasibility following these changes will be planned.

6.4 Justification for using an open-label design

EMPRESS will be conducted as an open-label trial as blinding the intervention is logistically impractical due to meropenem and piperacillin/tazobactam being dosed and administered differently and both having multiple important drug interactions that must be accounted for in clinical practice (43, 44).
In addition, the primary outcome (mortality) is an unequivocal outcome, which is unlikely to be affected by the clinician's knowledge of the intervention. This is supported by two previous meta-epidemiological studies that were unable to show firm associations between blinding and mortality in RCTs (45, 46); of note, another meta-epidemiological study found slightly larger intervention effects on mortality in unblinded trials but could not rule out confounding by other trial characteristics or small-study effects (47). Consequently, we consider the risk of substantial effects on the primary outcome and most secondary outcomes to be unlikely. Moreover, as blinding is costly and time-consuming for clinical personnel, we consider an open-label trial, which is capable of enrolling more patients at the same cost, more ethical, as it allows the final trial to be larger and thus have a higher chance of providing clinically conclusive results.
Due to the use of response-adaptive randomisation and regular adaptive analyses, the statistical team will

not be blinded; however, persons participating in data handling and statistical analyses will not be permitted to enrol patients or register outcome data.

6.5 Participant timeline

We aim to include participants as soon as possible after they fulfil all inclusion criteria and no exclusion criteria. The allocated treatment will be continued during stays in any participating site if the participant requires empirical antibiotics with either antibiotic assessed in the trial until a maximum of 30 days. When results from cultures and antimicrobial susceptibility tests are available, the treating clinician may change the antibiotics accordingly. Participants will be followed for 30 days after randomisation for most outcomes with additional longer-term outcomes collected after 90 and 180 days by use of electronic patient records, registers, and contacts to survivors. Where necessary (because participants cannot respond themselves), we will reach out to participants' next of kin to obtain data on health-related quality of life (HRQoL) outcomes. While HRQoL data obtained via proxies may not perfectly represent participants' HRQoL (48, 49), proxy HRQoL will only be used for participants that cannot respond themselves (between 2% and 13% of respondents in previous trials by our group (50-52)) and is preferable to having missing HRQoL data.

End of trial

The trial will stop inclusion of patients when one of the pre-defined stopping rules have been met, i.e., according to the adaptive analyses as described in section 12.4; following this, the trial will end when the last enrolled participant has completed 180-day follow-up (last-patient last-visit).

The end-of-trial will be reported no later than 90 days after the last-patient last-visit via CTIS. If earlier than planned, the reasons for stopping the trial will also be reported. The results of the trial will be reported on CTIS no later than 12 months after last-patient last-visit.

7 Selection of participants

Patients admitted to an active trial site will be considered for participation; patients will be eligible if they fulfil all inclusion criteria and no exclusion criteria outlined below; this will be presented in a flowchart as illustrated in **Figure S1** in appendix 3, section 19.3. Detailed definitions of inclusion and exclusion criteria are provided in appendix 4, section 19.4.

7.1 Inclusion criteria

All the following criteria must be fulfilled:

- Age ≥ 18 years
- Sepsis (including septic shock) defined according to the Sepsis-3 criteria (1), i.e., suspected or documented infection and an acute increase of ≥ 2 points in the Sequential Organ Failure Assessment (SOFA) score (a marker of acute organ dysfunction) (53)
- Critical illness defined as use of at least one of the following:
    a. Invasive mechanical ventilation
    b. Non-invasive ventilation
    c. Continuous use of continuous positive airway pressure (CPAP) for hypoxia
    d. Oxygen supplementation with an oxygen flow of ≥ 10 litres (L)/minute independent of delivery system and total flows
    e. Continuous infusion of any vasopressor or inotrope (excluding strictly procedure-related infusions)
- Clinical indication for empirical treatment with either meropenem or piperacillin/tazobactam

7.2 Exclusion criteria

We will exclude patients who fulfil any of the following criteria:

- Preceding intravenous treatment with meropenem or piperacillin/tazobactam for > 24 hours prior to screening
- Fertile women < 60 years of age with known pregnancy or positive urine human gonadotropin (hCG) or plasma hCG
- Known hypersensitivity or allergy to beta-lactam antibiotics
- Suspected or documented central nervous system infection

- Known infection/colonialization with microorganism with acquired resistance against meropenem or piperacillin/tazobactam within the previous 3 months (e.g., ESBL-, AmpC- or carbapenemase-producing bacteria)
- Current or planned use of valproate within 30 days from randomisation
- Patient included in another interventional trial where co-enrolment with EMPRESS is not permitted
- Previously randomised into the EMPRESS trial
- Informed consent following inclusion expected to be unobtainable
- Patient under coercive measures

7.3 Co-enrolment

We will generally allow and encourage co-enrolment with other interventional trials for the reasons outlined in appendix 5, section 19.5, except where the trial protocols collide. Co-enrolment agreements will be established with the sponsor/principal investigator of other relevant trials, and an updated list of trials approved for co-enrolment will be available to the participating sites and investigators.

7.4 Participant discontinuation and withdrawal

The procedure for handling withdrawal of consent from a participant will follow national regulations.
In Denmark, the procedure will be as outlined below:

Discontinuation and withdrawal at the choice of the participant or the proxy

A participant, who no longer wishes to participate in the trial, can withdraw their consent at any time without need of explanation and without consequences for further treatment. For incapacitated participants, consent can be withdrawn at any time by whoever has given proxy consent. To limit the amount of missing data, we will collect as many data as possible from each participant. Therefore, an investigator will ask the participant or the proxy if they allow continued data registration and follow-up up to day 180 despite discontinuation of the trial intervention.

Discontinuation at the choice of the investigator

A participant may have the intervention stopped by the clinician or investigator at any time, if:

- The participant experiences intolerable adverse events suspected to be related to the trial intervention (i.e., serious adverse reactions [SARs] or suspected unexpected serious adverse reactions [SUSARs]).

- The clinicians in conjunction with the coordinating investigator decide it is in the interest of the participant (e.g., due to potential interactions with other drugs that cannot be handled by dose-adjustment).
- The participant is withdrawn from active therapy.
- The participant is subject to coercive measures (e.g., ongoing involuntary hospital admission or under the jurisdiction of correctional authorities).

In these participants, the collection of data and the follow-up will continue, and the participant will remain in the intention-to-treat population.

Discharge

The trial allocation will be stopped when participants are discharged or transferred to a non-participating hospital department. The participant will still be followed through the electronic patient records. Participants who are discharged or transferred to a site participating in the EMPRESS trial will continue the allocated treatment at the new trial site if still relevant. If the participant is transferred or readmitted to an EMPRESS trial site from a non-participating site within 30 days of randomisation, the allocation will also resume if relevant.

8 Selection of trial sites and personnel

8.1 Trial sites and setting

Trial sites will be hospitals in Denmark and other countries, with a list of participating sites/countries continuously updated in the Clinical Trials Information System (CTIS).

8.2 Trial personnel

All clinical staff caring for patients will be eligible to care for and give the interventions to the trial participants.

The primary trial personnel are constituted of a dedicated team of research nurses, medical students, and clinical doctors, who are trained and certified in all trial-related procedures. The screening of patients in the electronic case report form (eCRF) will exclusively be done by the primary trial personnel.

The enrolment and informed consent procedure will follow the applicable legislation for clinical trials conducted in emergency situations (justification provided in appendix 6, section 19.6) and described separately for each participating country prior to start of enrolment in each country. The informed consent procedure will be handled by trained trial staff with thorough insight into the protocol.

After screening and randomisation of a participant, the trial staff or the clinical nurses will prepare the trial medication. The trained trial staff will also perform the necessary data entry. All participating trial sites will receive written and oral instructions about the trial procedures. A 24-hour telephone hotline will be available for trial-related questions.

8.3 Trial interventions

The intervention period is up to 30 days from randomisation or until discharge from a participating site, death, termination of empirical treatment including initiation of definitive antibiotic treatment, or de-escalation to another empirical antibiotic with more narrow spectrum (not including the two trial interventions), i.e., in situations where no bacteria are found in microbiological cultures, whichever comes first. We will collect data on the use of antibiotics on each day during the whole intervention period.

If test results from cultures and antimicrobial susceptibility tests show that de-escalation is not possible, but that there is presumed equal sensitivity to the two trial drugs, we recommend continuing the intervention as allocated.

When admitted or transferred to a participating site, the intervention must be resumed as allocated, if the participant has a new indication for empirical broad-spectrum antibiotic treatment (i.e., a new bacterial infection) within 30 days of randomisation. When admitted to non-participating departments, we recommend resuming the allocated intervention if the participant has a new indication for empirical broad-spectrum antibiotic treatment, but this decision will be at the discretion of the treating clinician.
From day 31 through 180, we will recommend using the same empirical antibiotic strategy as per the trial allocation if there is a new indication for empirical broad-spectrum antibiotic treatment, but we will not collect data on the use of antibiotics after the intervention period (i.e., day 30) and do not consider this part of the trial intervention.

8.4 Experimental intervention

The experimental intervention is IV meropenem 1 g given three times daily for up to 30 days until discharge from participating site, death, termination of empirical therapy (e.g., initiation of definitive antibiotic therapy), or de-escalation to another empirical antibiotic with more narrow spectrum (not including the two trial interventions), i.e., in situations where no bacteria are found in microbiological cultures, whichever comes first.

Preparation of experimental intervention: meropenem
Meropenem trihydrate comes in vials of 1 g and 0.5 g of meropenem. The experimental intervention will be prepared as below or according to local procedures.

Intermittent infusion
1 g of meropenem will be mixed with 20 mL of sterile water to a solution with a concentration of 50 mg/mL. The injection is administered over 5-30 minutes.

Prolonged or continuous infusion
1 g of meropenem will be mixed with 50-1000 mL of isotonic saline (0.9%) to a solution with a concentration of 1-20 mg/mL. The solution is administered as an infusion as per the clinical team (e.g., prolonged infusion over 3-4 hours or as continuous infusion).

Daily dose adjustments
Daily dose adjustments will be allowed at the discretion of the treating clinician, including but not limited to

the following examples. For participants with impaired kidney function (i.e., reduced glomerular filtration rate), the daily dose will be reduced at the discretion of the treating clinician. Suggested dose adjustments are provided in **Table 2**.

The daily dose may also be increased at the discretion of the treating clinician for participants with high creatinine clearance, difficult to treat infections, or a minimal inhibitory concentration presumably higher than normal. In these cases, we suggest administering meropenem 1 g x 4 as suggested in a study of Danish critically ill patients with sepsis in the ICU (54).

We will allow dose adjustments based on therapeutic drug monitoring and additional loading doses as per clinical practice at the participating sites.

Table 2: Suggested dose adjustments according to kidney function for meropenem (55, 56)

Glomerular filtration rate	**Dose and administration interval**
10-30 mL/min or continuous renal replacement therapy	1 g twice daily
0-10 mL/min	1 g once daily
Continuous renal replacement therapy	1 g twice daily
Intermittent haemodialysis	1 g once daily with a new dose given immediately after haemodialysis

8.5 Control intervention

The control intervention is IV piperacillin/tazobactam 4/0.5 g given four times daily for up to 30 days until discharge from participating site, death, termination of empirical therapy (e.g., initiation of definitive antibiotic therapy), or de-escalation to another empirical antibiotic with more narrow spectrum (not including the two trial interventions), i.e., in situations where no bacteria are found in microbiological cultures, whichever comes first.

Preparation of control intervention: piperacillin/tazobactam

Piperacillin sodium comes in vials of 4/0.5 g or 2.25 g of piperacillin/tazobactam. The control intervention will be prepared as below or according to local procedures.

Intermittent infusion

4/0.5 g of piperacillin/tazobactam will be mixed with 20 mL of sterile water to a solution with a concentration of 172.7 mg/mL of piperacillin and 22 mg/mL of tazobactam. The solution is shaken for 5-10 minutes until the powder has dissolved. The injection is administered over 5-30 minutes.

Prolonged or continuous infusion

4/0.5 g of piperacillin/tazobactam will be mixed with 20 mL of sterile water to a solution with a concentration of 172.7 mg/mL of piperacillin and 22 mg/mL of tazobactam. The solution is shaken for 5-10 minutes until the powder has dissolved. The solution is then mixed with isotonic saline (0.9%) to a total volume of 50-150 mL and a concentration of 26.7 to 80 mg of piperacillin and 3.3 to 10 mg of tazobactam. The solution is administered as prolonged infusion over 3-4 hours or as continuous infusion as per the clinical team.

Daily dose adjustments

Daily dose adjustments will be allowed at the discretion of the treating clinician, including but not limited to the following examples. For participants with impaired kidney function (i.e., reduced glomerular filtration rate), the daily dose will be reduced at the discretion of the treating clinician. Suggested dose adjustments are provided in **Table 3**.

For participants with high creatinine clearance, difficult to cure infections, or a minimal inhibitory concentration presumably higher than normal, the daily dose may be increased at the discretion of the treating clinician.

We will allow dose adjustments based on therapeutic drug monitoring and additional loading doses as per clinical practice at the participating sites.

Table 3: Dose adjustments according to kidney function for piperacillin/tazobactam (55, 56)

Glomerular filtration rate or haemodialysis	**Dose and administration interval**
10-30 mL/min or continuous renal replacement therapy	4/0.5 g three times daily
0-10 mL/min	4/0.5 g twice daily
Continuous renal replacement therapy (56)	4/0.5 g three times daily
Intermittent haemodialysis	4/0.5 g twice daily and an additional dose after every haemodialysis

8.6 Co-interventions

All participants in the trial will be given co-interventions at discretion of the treating clinicians. Of note, we will allow antibiotic combination therapy in both arms (e.g., empirical use of other antibiotics in addition to either trial drug except for piperacillin/tazobactam or meropenem, which may only be administered as allocated).

8.7 Monitoring of participants

The participant will be monitored closely due to the severity of their illness. The level of monitoring will be as per the clinical standard of the trial sites, often including measurements of oxygen saturation, heart rate, blood pressure, and respiratory rate; 8-hourly measurement of body temperature; and daily measurement of blood values, e.g., C-reactive protein (CRP), procalcitonin, leukocyte count, haemoglobin, creatinine, urea and electrolytes, pH, atrial blood gases, lactate, and blood glucose. Additional measurements will be done on clinical indications, including microbiological cultures, markers of candida infections and electrocardiograms (ECGs).

These data will not be registered in the EMPRESS trial eCRF but will be available in the participant's electronic health record for the investigators, monitors, and/or the authorities if needed.

8.8 Criteria for modification of interventions for a given trial participant

If protocol adherence is assessed by the clinical team to lead to suboptimal treatment of participants, the clinical team may at any time deviate from the protocol to ensure participant safety. Deviations will be registered as protocol violations according to the criteria outlined in section 8.9. We will have a 24-hour available EMPRESS trial hotline to enable discussion around-the-clock between the clinicians caring for trial participants and the EMPRESS trial team regarding protocol-related issues. Protocol violations will be registered and reported (sections 8.9 and 11.4).

8.9 Assessment of protocol adherence

We will record the occurrence of protocol violations on each day during the intervention period defined as:

- Use of empirical antibiotic treatment regimes, which do not include the allocated intervention, except when empirical antibiotic treatment is de-escalated to another empirical antibiotic with more narrow spectrum due to clinical improvement (not including the two trial interventions), i.e., in situations where no bacteria are found in microbiological cultures.
- Trial medication not administered per protocol defined as less than 50% of the prescribed daily dose of trial medication on at least one day during the intervention period.

Of note, combination therapy with other agents will not be considered a protocol violation. We will monitor protocol adherence at each trial site through the eCRF and alert trial sites in the case of clear violations (central monitoring). In addition, the trial will be externally monitored according to the Good Clinical Practice (GCP) directive and the monitoring plan (section 13).

8.10 Intervention accountability

Both trial interventions are routinely used for in-hospital treatment of patients with severe bacterial infections. In the EMPRESS trial, we will use shelf-medication provided by the participating sites' pharmacies. The trial medication will only be handled by trial staff, who are trained in the trial-related procedures, or by the clinical staff, who are trained and certified for the caring of these patients. The methods used for trial medication preparations are described in section 8.4 and 8.5.

Trial medications

Experimental intervention

Active medication

Meropenem, solution for injection and infusion, 1 g or 0.5 g of meropenem (as meropenem trihydrate) per vial, anatomical therapeutic chemical (ATC) code: J01DH02. Each 0.5 mg vial contains 1.96 mmol (i.e., 45.13 mg) of sodium and each 1 mg vial contains 3.92 mmol (i.e., 90.25 mg) of sodium.

Solution

Sterile water, solution for injection or infusion, ATC code: V07AB.

Isotonic saline, solution for intravenous injection, 9 mg/mL, ATC code: B05BB01. Each mL contains 9 mg of saline in sterile water. Content of electrolytes/L: 154 mmol chloride, 154 mmol natrium. Osmolarity 308 mmol/L.

We will allow the use of other isotonic fluids (e.g., isotonic glucose) at the discretion of the treating clinicians.

Control intervention

Active medication

Piperacillin, tazobactam, solution for injection and infusion, 4 g of piperacillin (as piperacillin sodium) and 0.5 of tazobactam (as tazobactam sodium) or 2 g of piperacillin (as sodium salt) and 0.25 of tazobactam (as tazobactam sodium) per vial, ATC code: J01CR05. Each 4/0.5 g vial contains 9.7 mmol (i.e., 224 mg) of sodium and each 2.25 g vial contains 4.9 mmol (i.e., 112 mg) of sodium.

Solution

Sterile water, solution for injection or infusion, ATC code: V07AB.

Isotonic saline, solution for intravenous injection, 9mg/mL, ATC code: B05BB01. Each mL contains 9 mg of saline in sterile water. Content of electrolytes/L: 154 mmol chloride, 154 mmol natrium. Osmolarity 308 mmol/L.

We will allow the use of other isotonic fluids (e.g., isotonic glucose) at the discretion of the treating clinicians.

9 Outcome measures

Outcome definitions are provided in appendix 4, section 19.4.

9.1 Clinical outcomes

Primary outcome

All-cause mortality at day 30 after randomisation.

Secondary outcomes

- Number of participants with one or more serious adverse reactions (SARs, defined as anaphylactic shock to IV piperacillin/tazobactam or meropenem, invasive fungal infection, pseudomembranous colitis, or toxic epidermal necrolysis) within 30 days of randomisation.
- Number of participants with new isolation precautions due to one or more resistant bacteria within 30 days of randomisation.
- Days alive without life support (i.e., invasive mechanical ventilation, circulatory support, or renal replacement therapy [including days in between intermittent renal replacement therapy]) from randomisation to day 30 and 90.
- Days alive and out of hospital from randomisation to day 30 and 90.
- All-cause mortality at day 90 and 180.
- HRQoL at day 180 using EQ-5D-5L index values (57).
- HRQoL at day 180 using EQ VAS (57).

9.2 Feasibility outcomes

The following feasibility outcomes will be assessed for the full population (i.e., not stratified by treatment allocation) after the data collection for all feasibility outcomes for the first 200 participants has been completed.

- Time to completion of feasibility phase
- Recruitment proportion
- Proportion of participants without consent to the use of data
- Protocol adherence
- Retention proportion

The feasibility outcomes will be assessed according to the criteria outlined in **Table 4**, which also describes the expected precision for all binary feasibility outcomes. If all feasibility outcomes fulfil the outlined criteria, the trial will continue unaltered; if not, modifications to the protocol will be applied (if necessary, after approval by the competent authorities, and if required with an inclusion pause until modifications are approved and implemented). The trial may also be stopped if otherwise deemed infeasible by the management committee and the independent data monitoring and safety committee (IDMSC).

Table 4: Feasibility criteria

	Feasible	**Expected precision**
Time to completion of enrolment for the feasibility phase	< 12 months	-
Recruitment proportion	≥ 50 %	43% to 57%
Proportion of participants without consent to the use of data	< 5%	2% to 9%
Protocol adherence	≥ 75%	68% to 81%
Retention proportion	≥ 95%	91% to 98%

For all binary feasibility outcomes, the expected precision of the proportions presented are the 95% confidence intervals (CIs; rounded) that would be obtained with 200 participants and a proportion matching the thresholds used to define the feasibility criteria calculated using the Clopper-Pearson method.

10 Safety

10.1 Definitions

We use the following definitions in accordance with the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) Harmonised Guideline for Good Clinical Practice (58):

Adverse event (AE)
Any undesirable medical event occurring to a participant during a clinical trial, which does not necessarily have a causal relationship with the intervention.

Adverse reaction (AR)
Any undesirable and unintended medical response related to the intervention that occurs at doses normally used in patients.

Serious adverse event (SAE)
Any adverse event that at any dose results in death, is life-threatening, requires hospitalisation or prolongation of existing hospitalisation, results in persistent or significant disability or incapacity, or is a congenital anomaly or birth defect.

Serious adverse reaction (SAR)
Any SAE assessed to be related to the intervention. In EMPRESS, expected SARs are defined as identified in the Danish Summary of Product Characteristics (SmPC) for piperacillin/tazobactam and meropenem.

Suspected unexpected serious adverse reaction (SUSAR)
Any suspected AR which is both serious and unexpected (the nature or severity of which is not consistent with the SmPC for piperacillin/tazobactam or meropenem).

10.2 Risk and safety issues in the EMPRESS trial

The trial participants will be hospitalised patients for whom adverse events and reactions are documented routinely in the electronic health record (i.e., notes and laboratory reports).

Risk adaptations to safety reporting in the EMPRESS trial

We consider the EMPRESS trial as a low-risk trial in regard to safety reporting (corresponding to risk level 1 in the Danish Medicines Agency's guidance on risk adaptations to adverse event recording (59)) in accordance with EU regulations (Regulation EU No 536/2014 (60)) as all of the following criteria are met:

1. Both interventions (i.e., meropenem and piperacillin/tazobactam) are marketed (and have been so for decades)
2. Both interventions are currently regarded as part of standard of care for patients with sepsis and septic shock
3. The safety profiles of both drugs are well-known, none of the drugs are subject to stricter reporting requirements in Denmark (61), and expectations of new safety issues are minimal

We will record and report the number of participants with specific SARs as well as SUSARs (see below). For SARs, we will report the number of participants with one or more SARs (defined as anaphylactic reaction to IV piperacillin/tazobactam or meropenem, invasive fungal infection, pseudomembranous colitis, and toxic epidermal necrolysis, section 9.1) from randomisation and up to day 30 for all participants, and report this as a secondary outcome in the primary trial report. Any SUSAR will be recorded and reported as described below.

As the EMPRESS trial is regarded as a low-risk trial, SAEs will not be recorded, however, the investigators will be able to report any SAEs immediately at their own discretion. Of note, we expect SAEs to be reflected in the other outcomes, including all-cause mortality, days live without life support, days alive and out of hospital, and HRQoL.

10.3 Assessment of serious adverse reactions

The following procedure for the assessment and reporting of SARs will be used in EMPRESS, with the possibility for local adaptations in specific countries if required (section 10.4).

Timing

In all participants, we will assess the occurrence of SARs and SUSARs in the 30 days following randomisation equal to the maximum intervention period.

Classification of an event

We will make no inferences about a causal relationship between the intervention and the expected SARs, as they are all known adverse reactions as per the SmPC. We thus *a priori* expect causality for all events defined as expected SARs. As the mechanisms of action for both interventions are comparable, the same expected SARs will be registered for both interventions. Thus, we will register the occurrence of events in the two groups and report them in the final report according to the definition given above. We have done this in previous critical care trials and expect that this procedure will result in complete and unbiased reporting given that the trial is open-label. An assessment of causality will be made by the investigators for all SUSARs.

Reporting

Any SUSAR (i.e., any suspected adverse reaction which is both serious and unexpected according to the investigator) will be reported within 24 hours to the sponsor or his delegate. If the SUSAR is adjudicated as fatal or life-threatening, the sponsor will report it to all trial sites as well as to the competent authorities in the participating countries within 7 days (for countries in EU via CTIS; for countries outside EU in collaboration with the national investigator). No later than 8 days after the reporting, the sponsor will inform the competent authorities via CTIS of relevant information on follow-up actions by the sponsor and the investigator. Any other SUSARs (i.e., those that are not life-threatening or fatal) will be reported via the EudraVigilance system to the competent authorities no later than 15 days from the time when the sponsor is informed.

The sponsor will submit an annual safety report, describing the rules of registration and reporting of safety events as described in this protocol, and including a list of all SARs, SUSARs, and investigator-reported events that have occurred at all sites during the trial period to the relevant entities through CTIS.

10.4 Country-specific adaptations of safety data

The registration and reporting of safety data in EMPRESS will follow EU regulations and be in accordance with the procedure outlined above (sections 10.2 and 10.3). Country-specific adaptations of safety data registration and reporting may be required for countries outside the EU. In these instances, the procedures

for registering and reporting additional safety data will be specified in the country-specific versions of the protocol.

10.5 Benefit-risk assessment

Both interventions are routinely used in clinical practice as part of standard treatment for critically ill patients with sepsis (6, 7) and are marketed in the EU, as described in detail in section 4.3. The safety profiles of both interventions are well-known, expectations of new safety issues are minimal, and the trial is assessed as being a low-risk trial as described in detail in section 10.2. Expected SARs based on the Danish SmPCs will be registered in both groups (section 9.1), included in the annual safety reports (section 10.3), and assessed yearly by the IDMSC along with any SUSARs (appendix 1, section 19.1). The IDMSC may recommend stopping the trial if any unexpected safety signal should occur. Consequently, the anticipated risks for participants are expected to be minimal and comparable to standard clinical practice. As use of piperacillin/tazobactam appears to be somewhat higher than use of meropenem (6, 7), the initial equal allocation to both interventions is expected to increase the proportion of participants receiving meropenem compared to standard practice. Based on low certainty evidence from the most recent systematic review and meta-analysis (22), this, if anything, is expected to lead to overall better outcomes for participants versus comparable patients not in the trial, as reflected by the trial hypothesis. The use of response-adaptive randomisation ensures that more participants will be allocated to the intervention with the highest probability of being superior following the first adaptive analysis, which also constitutes a direct benefit to participants. In summary, the anticipated benefits of trial participation are assessed to outweigh the corresponding minimal risks.

11 Procedures, assessments, and data collection

11.1 Screening

Adult patients who fulfil all inclusion criteria and are admitted to a participating trial site will be eligible for screening.

The screening will be done by trained trial staff (trial investigators) assessing the eligibility criteria (section 7.1 and section 7.2). The treating clinical doctor may always decline participation on behalf of a patient fulfilling the eligibility criteria if he/she consider it to be in the best interest of the patient.

For all fertile women under 60 years of age, screening for human chorionic gonadotropin (hCG) in urine or plasma will be done before enrolment in the trial. If a hCG-test has already been done under the current admission, we will use the test result of this for screening for pregnancy.

11.2 Procedures of informed consent

Procedures for enrolment and informed consent will follow the applicable legislation for clinical trials conducted in emergency situations in the participating countries, as critical illness due to sepsis constitutes an emergency situation requiring immediate empirical broad-spectrum antibiotics, as per clinical guidelines (62). A detailed justification of this is provided in appendix 6, section 19.6. Consequently, enrolment will be without prior informed consent from the patients. Details on the recruitment and informed consent procedure in each participating country will be available on the trial website prior to start of enrolment in each participating country.

11.3 Data collection

The screening of participants in the eCRF will be done by trained trial staff as described in section 11.1. After screening and randomisation, the data below (section 11.4) will be obtained and entered in the eCRF by the trial staff using data from the electronic patient records, healthcare registers and interviews with participants or relatives. For participants transferred from a trial site to a non-trial site, outcome data will be collected from the electronic patient records, healthcare registers, or by contact to staff at the non-trial site.

11.4 Variables

All variables are defined in appendix 4, section 19.4. Mock baseline and outcome tables are presented in **Tables S1 and S2** appendix 7, section 19.7.

Screening variables

- Inclusion and exclusion criteria (sections 7.1 and 7.2)
- Age at enrolment (date of birth)
- Presence of haematological or metastatic cancer (yes/no)
- Acute surgical admission (yes/no)
- Use of respiratory support at randomisation (yes/no)
    - Invasive mechanical ventilation
    - Non-invasive ventilation
    - Continuous use of CPAP
    - Supplementary oxygen with an oxygen flow ≥10 L/min
- Latest fraction of inspired oxygen (FiO_2) prior to randomisation (only for participants on invasive mechanical ventilation, non-invasive ventilation, or continuous use of CPAP)
- Maximum oxygen flow at randomisation (+/- 1 hour, only for participants receiving supplementary oxygen on an open system)
- Use of circulatory support (infusion of vasopressor/inotropes) at randomisation (yes/no)
- Use of renal replacement therapy within the last 72 hours prior to randomisation (yes/no)
- Trial site

Baseline variables

- Sex
    - Male
    - Female
- Coexisting conditions
    - History of ischaemic heart disease or heart failure (yes/no)
    - Diabetes mellitus (yes/no)
    - Chronic pulmonary disease (yes/no)
    - Known use of immunosuppressive therapy within the last 3 months (yes/no)
    - Previous organ transplantation (yes/no)
    - Chronic liver disease (yes/no)

- Chronic use of renal replacement therapy (yes/no)

- Clinical Frailty Scale, version 2.0 (level 1-9) (63, 64)

- Date of admission to hospital

- Type of department at which the participant was included
    - Emergency department
    - Hospital ward
    - Intermediate care unit
    - ICU

- Primary site of infection (suspected or confirmed)
    - Pulmonary
    - Gastrointestinal
    - Urinary tract
    - Skin or soft tissue
    - Bloodstream
    - Other, including unknown primary site of infection

- Positive bacterial cultures from usually sterile sites taken in the 48 hours prior to and 1 hour after randomisation (yes/no)
    - If yes, apply type of sample
        - Blood
        - Lower airway secretions or bronchoalveolar lavage
        - Urine
        - Tissue, bone, or pus
        - Peritoneal fluid
        - Pleural fluid
        - Cerebrospinal fluid
    - For each type of sample, apply type of bacteria
        - *Achromobacter* spp.
        - *Acinetobacter baumanii*
        - *Aerococcus* spp.
        - *Aeromonas* spp.
        - Anaerobes
        - *Burkholderia* spp.
        - *Campylobacter* spp.

- *Citrobacter* spp.
- *Enterobacter* spp.
- *Enterococcus faecalis*
- *Enterococcus faecium*
- Other *Enterococcus*
- *Escherichia coli*
- *Haemophilus influenza*
- *Klebsiella* spp.
- *Legionella* spp.
- *Listeria* spp.
- *Moraxella catarrhalis*
- *Neisseria* spp.
- *Proteus* spp.
- *Pseudomonas* spp.
- *Salmonella* spp.
- *Serratia* spp.
- *Shigella* spp.
- Methicillin-resistant *Staphylococcus aureus (*MRSA)
- Methicillin-sensitive *Staphylococcus aureus (*MSSA)
- *Stenotrophomonas* spp.
- *Streptococcus pneumoniae*
- Other *Streptococcus*
- Other bacteria

- For each type of bacteria
  - Resistance to piperacillin/tazobactam (yes/no)
  - Resistance to meropenem (yes/no)

- Limitations of care (i.e., not candidate for invasive mechanical ventilation, circulatory support, renal replacement therapy, cardio-pulmonary resuscitation) at the time of randomisation (yes/no)
- Treatment during current hospital admission prior to randomisation:

- Anti-bacterial agent (IV, oral, or per gastrointestinal tube) (yes/no), if yes:
  - Beta-lactam/beta-lactamase inhibitor
  - Carbapenem
  - Cephalosporin

    - Penicillin
    - Glycopeptide
    - Fluoroquinolone
    - Macrolide
    - Clindamycin
    - Nitroimidazole
    - Aminoglycoside
    - Oxazolidinone
    - Sulfonamide
    - Tetracycline
    - Other

- Body weight, vital parameters, and blood values:
    - Participant weight (kg) (estimated or measured)
    - Participant height (m) (estimated or measured)
    - Partial pressure of arterial oxygen (PaO_2) and arterial oxygen saturation (SaO_2) in the most recent arterial blood gas sample prior to inclusion **OR** peripheral oxygen saturation (SpO_2) from pulse oximeter if arterial blood gas sample is not available
    - Lowest registered systolic blood pressure (mmHg) in the 24 hours preceding randomisation
    - Highest plasma lactate in the last 24 hours prior to randomisation (mmol/L)
    - Highest plasma creatinine in the 24 hours prior to randomisation (µmol/L)

Daily during admission for the first 30 days after randomisation (day forms)

- Death (yes/no)
- Admitted to hospital (yes/no)
- Use of invasive mechanical ventilation (yes/no)
- Use of circulatory support (continuous infusion of vasopressor/inotropes for a minimum of 1 hour) on this day (yes/no)
- Use of any form of renal replacement therapy on this day including days between intermittent renal replacement therapy (yes/no)
- SAR(s) on this day (yes/no for each)
    - Anaphylactic shock as a reaction to IV piperacillin/tazobactam or meropenem
    - Invasive fungal infection
    - Pseudomembranous colitis

- Toxic epidermal necrolysis

- New isolation precautions due to resistant bacteria on this day (yes/no)
  - If yes, type of resistant bacteria:
    - Vancomycin-resistant *Enterococci*
    - *Clostridiodies difficile*
    - Carbapenemase-producing bacteria
    - Extended spectrum β-lactamase (ESBL) producing *Enterobacterales*
    - AmpC β-lactamase-producing *Enterobacterales*
    - *Citrobacter freundii*
    - Methicillin-resistant *Staphylococcus aureus* (MRSA)
    - *Stenotrophomonas maltophilia*
    - Other

Daily registration of co-interventions and protocol adherence up to day 30 after randomisation (only during intervention period)

-Trial intervention: did the participant receive at least 50% of the prescribed daily dose of trial medication on each day (yes/no)

- If no, apply reason:
  - Change to definitive treatment
  - De-escalation to another empirical antibiotic with more narrow spectrum due to clinical improvement (not including the two trial interventions), i.e., in situations where no bacteria are found in microbiological cultures
  - Empirical or definitive antibiotic treatment no longer indicated (including switching to prophylactic antibiotic treatment)
  - Discharged to non-participating department
  - Clinical deterioration with indication for different antibiotic than allocated according to treating clinician
  - Adverse reaction to the allocated agent requiring shift of antibiotic according to treating clinician
  - By error
  - Other reason(s)

- Use of any other antibiotic agent than allocated on this day (yes/no)
  - If yes, apply type of treatment

    - Prophylactic treatment
    - Empirical treatment
    - Definitive treatment
- If yes, apply type of agent
    - Beta-lactam/beta-lactamase inhibitor
    - Carbapenem
    - Cephalosporin
    - Penicillin
    - Glycopeptide
    - Fluoroquinolone
    - Macrolide
    - Clindamycin
    - Nitroimidazole
    - Aminoglycoside
    - Oxazolidinone
    - Sulfonamide
    - Tetracycline
    - Polymyxin
    - Other

Daily during admission for day 31-90 after randomisation (day forms)

- Death (yes/no)

- Admitted to hospital (yes/no)

- Use of invasive mechanical ventilation (yes/no)

- Use of circulatory support (continuous infusion of vasopressor/inotropes for a minimum of 1 hour) on this day (yes/no)

- Use of any form of renal replacement therapy on this day including days between intermittent renal replacement therapy (yes/no)

Follow-up 180 days after randomisation

- Death (yes/no)

- If yes, date of death

- HRQoL assessed by EQ-5D-5L domains (level 1, 2, 3, 4 or 5 for each domain)

- HRQoL assessed by EQ VAS (0-100)

- Date of HRQoL assessment

    - For non-responders, provide reason(s) for not responding
        - Do not wish to participate
        - Death after day 180 but before HRQoL follow up was conducted
        - Contact to participant or proxy could not be established
        - Other

- HRQoL assessor

    - Participant
    - Proxy

12 Statistical analyses and adaptation

12.1 General considerations

EMPRESS is a Bayesian adaptive trial with an integrated feasibility phase (sections 6.2 and 9.2 with feasibility outcomes and trial adaptations/stopping based on feasibility criteria) employing adaptive stopping rules (for superiority/inferiority, practical equivalence, and at a maximum pre-specified sample size) and response-adaptive randomisation (25). The adaptation rules (including the simulations used to evaluate these rules) and statistical analyses plan are described in this section of the protocol, which constitutes the full statistical analysis plan. All statistical analyses are planned and will be conducted according to the ICH-GCP guidelines (65). The adaptive and Bayesian aspects have been planned and evaluated using statistical simulation (25) and in accordance with guidance from the Accelerating Clinical Trials in the EU (ACT-EU) Steering Group under the European Medicines Agency (EMA) (66) and the United States Food and Drug Administration (FDA) (67). This protocol and statistical analysis plan will be made publicly available online on the trial website before trial initiation, and we aim to publish the protocol in a peer-reviewed journal prior the first adaptive analysis.

Rationale for using Bayesian statistical methods

EMPRESS will solely use Bayesian statistical methods for both adaptive (interim) analyses and the final analyses. Bayesian statistical methods are common in adaptive trials and other advanced trial designs, as they are well suited for contexts where analyses are frequently updated based on accumulating evidence (25, 68-70). Compared to the more commonly used frequentist statistical methods, Bayesian methods have interpretational advantages. Frequentist statistical methods produce P-values and confidence intervals (CIs), both of which are difficult to understand and frequently misinterpreted (71-73). This is especially problematic for P-values, which are *indirect* probabilities related to the hypothesis investigated (i.e., that there is a difference in outcomes between the treatments assessed) calculated under the assumption that there is *exactly* no difference between treatments (69). In contrast, Bayesian statistical methods produce *direct* probabilities related to the effect sizes of interest and measures of uncertainty (credible intervals, CrIs) that are more intuitive to interpret than conventional, frequentist (CIs) (69). Bayesian statistical methods use probability distribution to express uncertainty and start with a *prior* probability distribution that expresses the analysts' belief in the treatment effect before conducting the trial. These *prior* probability distributions are updated with the trial data to yield *posterior* probability distributions, which are compromises between the prior belief in treatment effect and the observed trial data. In EMPRESS, all

primary analyses will use neutral weakly informative priors (described in section 12.7 and appendix 8, section 19.8) as mostly done in Bayesian analyses of trials conducted in critically ill patients (69). Neutral priors are centred on *no difference* (i.e., they do not favour one treatment arm over another) and weakly informative means that they do not contribute with a lot of information to the analysis compared to the actual number of participants included (i.e., they have minimal influence on the results), while they rule out *a priori* implausibly large effect sizes. Using weak priors encompassing slight scepticism regarding large treatment effects is sensible given that most trials conducted in critically ill patients are inconclusive or show relatively small effect sizes (74). The use of sceptical weakly informative priors in adaptive trials limits the risk of random erroneous conclusions and overly aggressive adaptation to chance findings early in the trial with limited influence on total sample sizes and event counts (75).

Rationale for using an adaptive trial design

Most RCTs conducted in critically ill patients use fixed sample sizes with no or very limited adaptation (i.e., usually few or no interim analyses prior to reaching the pre-specified sample size) (69). Such trials are inflexible and often inconclusive due to incorrect assumed baseline event probabilities and overly optimistic treatment effects in sample size calculations (74, 76-78). This often leads to results that are not *statistically significant* and erroneously interpreted as if there were no difference between treatment arms (69, 72, 73, 79). Misinterpretation of inconclusive trial results poses a risk of research waste and erroneous, premature abandoning of effective treatments, which may ultimately lead to harm or delayed improvements in patient care (25, 69). While assumptions about event probabilities and effect sizes are used when planning adaptive trials, trials with adaptive stopping are more likely to produce conclusive results, even if these assumptions do not hold, as they may continue past the expected sample size until a stopping rule has been triggered (25). Importantly, adaptive trials can be designed to have similar error rates as conventional non-adaptive or group sequential trial designs (25). Consequently, EMPRESS has been designed as a Bayesian adaptive trial to maximise the likelihood of conclusive results that will guide clinical practice and improve patient care.

12.2 Analysis populations

All adaptive (interim) analyses and the final analyses of EMPRESS will be conducted in all randomised participants for whom there is consent to the use of any data (intention-to-treat population). In addition, the final analyses of all outcomes will also be repeated in the *per-protocol* population, defined as those without any major protocol violations, as defined in section 8.9, and in the population of participants without baseline cultures with bacteria with known resistance to the allocated treatment.

12.3 Timing of adaptive analyses and implementation of adaptations

The first adaptive (interim) analysis will be conducted after an initial *burn-in* period during which 400 participants complete follow-up for the primary outcome (30 days after randomisation) and the data collection period (15 days after end of follow up, 45 days after randomisation). This is followed by adaptive analyses after each additional 300 participants have completed follow-up for the primary outcome until one of the pre-specified stopping rules (section 12.4) is triggered. If relevant, adaptive analyses will be conducted at the same intervals after inclusion of the maximum number of participants have been randomised but have not completed the outcome-data collection period yet. Adaptive analyses will take place on the first workday of each month or as soon as possible thereafter, if the number of participants having completed the combined follow-up and data collection period for the primary outcome (30 days follow-up plus 15 days data collection) reaches or surpasses any of these thresholds based on the number of participants. The choice to conduct adaptive analyses as close to specific dates as possible is pragmatic to make timely data collection and validation across sites feasible and synchronised. No adaptive analyses will take place until a final decision regarding feasibility (sections 6.2 and 9.2) has been made, which is expected to occur before the first planned adaptive analysis (i.e., after 400 participants have completed follow-up for the primary outcome). In the unlikely event that this is not the case, the first adaptive analysis will be conducted as soon as possible after the first workday of the next month after that. Only participants who have completed the combined follow-up and data collection period will be included in the adaptive analyses; participants who have not reached this time point will not be included, even if outcome data are available to avoid upwards bias of event probabilities due to inclusion of early deaths.

Following adaptive analyses, the IDMSC will be informed of the results, with adaptations implemented 48 hours after informing the IDMSC (or as soon as possible after) except in cases where the IDMSC objects to their immediate implementation (appendix 1, section 19.1).

12.4 Stopping rules

The pre-specified adaptive stopping rules in EMPRESS have been determined and evaluated using statistical simulations (appendix 9, section 19.9) (25, 80) with stopping rules calibrated to ensure a type 1 error rate for the primary outcome of 5%. In addition to the adaptive stopping rules, EMPRESS may be stopped after the feasibility phase (section 6.3 and 9.2) before any adaptive analyses if the trial is deemed infeasible.

EMPRESS will use *constant* stopping rules at all adaptive analyses. Compared to varying, increasingly lenient stopping rules (as frequently used for interim analyses in conventional group sequential trial designs (81), where early interim analyses use stricter thresholds than later analyses), constant rules increase the

probability of stopping for superiority earlier and thus decreases the expected (mean) sample sizes. It may further lower the potential overestimation of treatment effects that may occur if trials be stopped early, at the cost of a lower probability of stopping the trial for superiority at the final possible analysis (82-84). We expect EMPRESS to be stopped substantially earlier than the maximum allowed sample size, which is viewed as a *worst-case scenario* (section 12.6 and appendix 9, section 19.9). Consequently, the use of constant, as opposed to varying and decreasing, stopping rules fits the design of EMPRESS best.

Superiority/inferiority

The trial will be stopped for *superiority* if the probability that one arm is better than the other (regardless of how much) exceeds 99.64% in any adaptive analysis (and *vice versa* for *inferiority*, i.e., a probability of an arm being the best below 0.36%, as the other arm, by definition, is inferior if one arm is superior in a two-arm trial). Superiority will be assessed using the posterior distribution from the adaptive analyses of the primary outcome as described in section 12.7.

Practical equivalence

The trial will be stopped for *practical equivalence* if the probability that the absolute difference in the primary outcome of 30-day mortality between two arms is less than 2.5%-points in either direction exceeds 90% in any adaptive analysis. Practical equivalence will be assessed using the posterior distribution from the primary analysis of the primary outcome as described in section 12.7. Practical equivalence will always be assessed *after* superiority, so in the unlikely event that both criteria are fulfilled (i.e., the probability that one treatment is superior passes the threshold, while the probability threshold for both treatments being practically equivalent is also passed, which may occur if a treatment *is* truly superior, but the difference is small), the trial will be considered stopped for superiority as this is a more clinically useful decision.

Futility

No stopping rule for futility is used in EMPRESS, as stopping for futility is less clinically useful than equivalence and as both treatments assessed are already widely used.

Maximum sample size

The trial will enrol a maximum of 14,000 participants (due to logistic and economical concerns) with a final analysis conducted as soon as all these participants have completed the follow-up and data collection period. If this final analysis is conducted after follow-up and data collection for the maximum number of participants, the same criteria for *superiority* and *practical equivalence* as outlined above will be used; if

none of these are reached, the trial will be considered *inconclusive* regarding the effect sizes used for the assessment of *practical equivalence*, but the results will nonetheless be reported probabilistically.

Stopping, follow-up, final analysis, and communication of results

If the stopping rule for either *superiority/inferiority* or *practical equivalence* is triggered, the trial will stop randomising participants, while continuing follow-up for the remaining included participants. The final analysis will then be conducted once all participants complete the data collection period and data are available for all participants. The stopping-rule-based final decision will be reported as the conclusion of the trial, even if the stopping criteria are no longer fulfilled after the final analysis (due to random fluctuations when follow-up concludes for all randomised participants, final probabilities may differ slightly from the probabilities at the time of stopping); the final numerical results reported will be those from the final analysis including all participants with the results from the analysis leading to stopping secondarily reported.

Results from all adaptive analyses (and allocation ratios following each adaptive analysis) will be reported in a supplement to the primary report.

The stopping and the reason for stopping will be communicated as soon as possible to all participating sites. If stopped for superiority, the superior treatment arm will also be communicated to participating sites; the investigators and treating clinicians may then determine if participants still active in the trial should be switched to this treatment regardless of allocation. If stopped for practical equivalence or at the maximum sample size, the treatment of all active trial participants should continue according to their allocation. The stopping and stopping reason will be communicated publicly as soon as possible after stopping with complete, numerical trial results reported in full as described in section 12.7 within reasonable time.

12.5 Response-adaptive randomisation

As stated in section 6.2 (along with the rationale), we will initially use equal (1:1) allocation using stratified (by trial site) block randomisation (with randomly varying block sizes of 2, 4 or 6) according to computer-generated lists (39) until the first adaptive analysis. Once results from the first adaptive analysis are available, we will switch to simple (unstratified) response-adaptive randomisation (25). Response-adaptive randomisation will be based on the probabilities of each arm being superior (calculated as described in section 12.7 using the posterior distribution for the average treatment effect from the primary analysis of the primary outcome), which in a two-arm trial by definition is equal to the probability of the other arm

being inferior and *vice versa*. We will restrict the response-adaptive randomisation to allow minimum 40% and maximum 60% allocation to either group. Due to these restrictions, we will not use a *softening factor* to further restrict allocation probabilities (25).

12.6 Performance metrics (including expected sample size)

Performance metric evaluation, assumptions, and clinical scenarios

The final trial design has been developed and evaluated using statistical simulation as recommended (25, 66, 67, 80) and described in further detail in appendix 9, section 19.9. In brief, the final trial design was evaluated using 100,000 simulations as recommended (67) under multiple different clinical scenarios to assess the Bayesian analogues of the type 1 error rate for the primary outcome (i.e., the probability of stopping for superiority if there truly is no difference between arms) and the power (100% - the type 2 error rate, i.e., the probability of stopping for superiority if there truly is a difference between arms) (25). We assessed multiple performance metrics under three different scenarios for the primary outcome of 30-day mortality:

- *No difference/null scenario*: assuming identical event probabilities of 25% in both groups
- *Small difference*: assuming event probabilities of 25% in the piperacillin/tazobactam group and 22.5% in the meropenem group (corresponding to a relative risk of 0.90 and a risk difference of -2.5% with meropenem)
- *Large difference:* assuming event probabilities of 25% in the piperacillin/tazobactam group and 20% in the meropenem group (corresponding to a relative risk of 0.80 and a risk difference of -5% with meropenem)

In addition, we assumed inclusion rates corresponding to 5 participants/day throughout the trial (while they are expected to be lower during the feasibility phase, this was disregarded here, as the first adaptive analysis is expected to take place some time after completing the feasibility phase) and combined follow-up, data collection, and data verification lags of 45 days (85). All trial designs considered and compared were calibrated to ensure type 1 error rates for the primary outcome of approximately and maximally 5%. Simulations were largely similar with actual adaptive analyses, although some necessary simplifications (appendix 9, section 19.9) were used, e.g., no simulation or handling of adjustment variables in analyses in simulations; consequently, the actual analyses of the trial may have slightly higher power and require slightly lower sample sizes.

Multiple operating characteristics/performance metrics were considered during the simulations and trial design phase (86). In brief, both metrics important for practical/economical/logistical reasons (e.g., expected total sample sizes), benefit for *internal* patients (i.e., participants included in the trial, including total event probabilities and probabilities of being allocated to the best arm), benefit to *external* patients (i.e., those not included in the trial and future patients, e.g., the probabilities of conclusiveness/power/type 1 and 2 error rates, and ideal design percentages [IDPs]), and accuracy (root mean squared errors [RMSEs] of the effect estimated in the selected arm, if any) (25, 86). Some of these metrics (RMSE, IDPs) are calculated using two selection strategies (25): i) for simulations ending in superiority only and ii) assuming that the control arm (piperacillin/tazobactam) is selected in trials not ending in superiority, as this may best reflect clinical practice if the trial ends up with an equivalence decision (or inconclusive due to stop at a maximum sample size). Details on the performance metrics assessed and their definitions are included in appendix 9, section 19.9.

Performance metric evaluation and clinical scenarios

Key performance metrics under the three different scenarios assessed with the final, calibrated stopping rules (rounded to 4 significant digits) are presented here; all performance metrics assessed are presented for the final trial design and with the primary assumptions and in sensitivity analyses varying key assumptions, along with metrics for several design variants in appendix 9, section 19.9 (**Table 5** below is based on an excerpt from **Table S5** in the appendix):

Table 5: Key performance metrics of the final EMPRESS design under the primary assumptions used

Metric	No difference	Small difference	Large difference
Sample size – mean [25; 50; 75%-percentiles]	5189 [3925; 4525; 6025]	5859 [3625; 5125; 7825]	2570 [1525; 2425; 3325]
Event count – mean [25; 50; 75%-percentiles]	1297 [961; 1120; 1485]	1380 [852; 1227; 1855]	569 [336; 507; 739]
Event probabilities – mean [25; 50; 75%-percentiles]	25.0% [24.6%; 25.0%; 25.4%]	23.6% [23.2%; 23.6%; 24.0%]	22.2% [21.6%; 22.2%; 22.8%]
Probability of conclusiveness	99.7%	99.0%	100.0%
Probability of superiority	4.99%	72.3%	99.8%
Probability of equivalence	94.7%	26.7%	0.2%
Probability of inconclusiveness	0.3%	1.0%	0.0%
Root mean squared error (for estimated event probability in the superior arm, for simulations stopped for superiority only)	2.4 %-points	1.1 %-points	1.2 %-points
Root mean squared error for the treatment difference with meropenem (for simulations stopped with a superiority decision for meropenem only)	4.8 %-points	1.9 %-points	1.9 %-points

Key performance metrics of the final EMPRESS design under the primary assumptions used based on 100,000 simulations for each scenario. The probability of conclusiveness corresponds to the combined probabilities of superiority or equivalence; the probability of superiority may be interpreted as the type 1 error rate in the scenario with no difference and as the power in the scenarios with differences present (25). The probability of inconclusiveness refers to the proportion of simulated trials stopped at the maximum sample size without triggering a superiority or equivalence stopping rule.

12.7 Statistical analyses and models

Statistical analyses (including the adaptive analyses) are described below.

Population, statistical framework, and priors used

The primary statistical analyses will be conducted in the intention-to-treat population with secondary analyses in the per-protocol population, both described in section 12.2. All analyses will use Bayesian statistical methods and neutral, weakly informative, sceptical priors, i.e., symmetrical priors centred on no difference for the intervention effect. Such priors do not favour one treatment over another but put low prior probability on very large treatment effects, stabilises computations, and decreases the risk and magnitudes of errors and over-aggressive adaptation to chance early in an adaptive trial, while having limited influence on total sample sizes (75). Neutral, very weakly informative priors will be used for all additional model terms; these priors will have minimal influence on the estimation and results, but are preferable to completely uninformative (i.e., flat) priors as they stabilise the Bayesian models during fitting

and lead to faster model convergence (87). Exact prior specifications along with justifications and description of how much information each prior corresponds to, in understandable terms (i.e., the number of corresponding participants (75, 88)), are included in appendix 8, section 19.8. Sensitivity analyses using more sceptical priors and evidence-based priors (if relevant external evidence becomes available) for the treatment effect (and the same priors for all adjustment variables as in the primary analyses) will be conducted for the final analyses; these exact priors are also described in appendix 8, section 19.8.

General principles

Data will be presented descriptively stratified by treatment allocation for all variables outlined in section 11.4 using medians with 25% and 75% percentiles (interquartile ranges, IQRs) for numeric variables and numbers with percentages for categorical variables (**Tables S1 and S2**, appendix 7, section 19.7). Baseline variables included in **Table S1** (appendix 7, section 19.7) will additionally be presented separately for each time period (with a new time period defined each time allocation ratios are updated after an adaptive analysis) for both treatment groups combined and each treatment group separately in supplementary tables.

All outcomes will be analysed using adjusted (generalized) linear models, including the treatment allocation and all adjustment variables outlined below, with results primarily presented as average treatment effects calculated by a G-computation-like approach as previously described (89-91). In brief, effects will be standardised by marginalising over all covariates in the actual and counterfactual treatment assignment scenarios by first predicting the expected values for each participant under the actual and counterfactual treatment assignment scenarios, calculating the predictions in each group, and finally calculating the absolute and relative average treatment effects from these values (91). The entire posterior distributions will be used to appropriately propagate uncertainty through the G-computation procedure. Binary outcomes will be analysed using logistic regression models with adjusted risk differences (RDs) and risk ratios (RRs) calculated from expected probabilities under the factual and counterfactual scenarios regarding allocated treatment; the underlying, conditional/adjusted odds ratios (OR) for the final analyses will be presented secondarily. Numerical outcomes will be analysed using linear regression models with adjusted mean differences (MDs) and ratios of means (RoMs) calculated from the expected values under the factual and counterfactual scenarios regarding allocated treatment. Complete posteriors for the intervention effects will be presented visually (as densities and cumulative densities) and summarised numerically using median posterior values as point estimates with 95% percentile-based credible intervals. Probabilities of any benefit of meropenem treatment will be calculated on the absolute effect scale (RD or MD) using the posterior distribution of average treatment effects. In addition, probabilities of differences being smaller

than the threshold for practical equivalence (absolute risk difference of 2.5%-points) or larger in both the beneficial and harmful directions will be calculated and presented for the primary outcome. These probabilities on the absolute effect scale for the primary outcome will be used to guide all adaptations outlined above. This approach is both robust to potential model misspecification and appropriately handles eventual imbalances between treatment groups in the covariates adjusted for (91). We will use linear models to assess multiple outcomes with non-normal distributions with substantial inflation at specific values (e.g., days alive without life support with inflation at the minimum and maximum values); although these models are not restricted to the valid outcome space and may not perfectly replicate the data if used generatively, they adequately and robustly estimate the mean values in each group, which are used for all further calculations in the primary analyses (92). For composite outcomes, the individual components will secondarily be reported descriptively only.

Adjustment variables

All analyses will include the treatment indicator (piperacillin/tazobactam being the reference) and further adjusted for the following variables registered at the time of enrolment, as defined in section 11.4 with further details in appendix 4, section 19.4:

- Age (years, integer)
- Presence of haematological or metastatic cancer (yes/no)
- Acute surgical admission (yes/no)
- Use of invasive mechanical ventilation (yes/no)
- Use of circulatory support (continuous infusion of vasopressors/inotropes) (yes/no)
- Use of renal replacement therapy within the last 72 hours prior to randomisation (yes/no)
- Site of inclusion (categorical, with the largest site being the reference)
- Time period (categorical, with a new period defined each time allocation ratios are updated after an adaptive analysis, with the most recent period being the reference)

All binary/categorical variables will be modelled as simple additive effects on the appropriate scale (i.e., the log-odds scale for the logistic regression models and the actual scale for the linear regression models). To account for potential non-linearity with age (the only continuous variable adjusted for), both age and a quadratic transformation of age will be entered in the models. As the G-computation-based standardisation approach outlined above is robust to potential model misspecification (91), we believe this method of handling age strikes a good balance between not assuming linearity while also not introducing unnecessary complexity into the models. If necessary, sites with very few participants may be merged in the models if

otherwise causing convergence issues. Where relevant and possible, 'small' sites will preferably be merged with comparable sites within the same region; sites merged in earlier adaptive analyses will be included separately at later analyses if possible due to inclusion of more participants.

Handling of days alive without life support and out of hospital outcomes

The primary analyses of days alive without life support and days alive and out of hospital at day 30 and 90 will be conducted after assigning non-survivors the value 0. The rationale for this is that death is then considered the worst possible outcome in the analyses of these outcomes, respecting the overall outcome prioritisation in EMPRESS, and corresponding to the most common definitions and usage of these outcomes (93-95). In addition, sensitivity analyses will be conducted using the *actual* values without penalising death.

Handling of health-related quality of life outcomes

The primary analyses of the health-related quality-of-life outcomes (i.e., EQ-5D-5L index values and EQ VAS) will be conducted in all participants; this will be supplemented with secondary analyses conducted in survivors only. For both outcomes, non-survivors will be assigned the value 0, which, by definition, corresponds to a health state as bad as being dead for EQ-5D-5L index values and is the worst possible value for EQ VAS (57). EQ-5D-5L index values in survivors will be calculated using national value sets where available as recommended (96), using the value set adjudicated as most appropriate for countries without their own national value sets (determined by the management committee in correspondence with the national investigator in each participating country and using the Danish value set (97) if no other good option is available, as most participants are expected to be enrolled in Denmark). A sensitivity analysis will be conducted using the Danish value set for all participants (as most participants are expected to be enrolled in Demark) as recommended (96). The raw numbers and proportions of responses in each domain will be presented for respondents only (i.e., complete cases).

Model fitting, assessment of convergence, and reporting

Models will be fit using the default dynamic Hamiltonian Monte Carlo sampler in Stan (98) through the *brms* (99) R package with assessment of model adequacy performed as previously described (90, 100). We will use 4 chains with a total minimum of at least 30,000 post-warm-up samples and require bulk/tail effective sample sizes of at least 1,000 for all parameters and at least 10,000 for the treatment allocation parameter. We will tune sampler settings as necessary to avoid divergent transitions and evaluate chain convergence using the updated *Rhat* statistics (required to be ≤1.01 for all parameters) (101) and visual

inspection of overlain trace and density plots (102, 103) increasing the number of samples as necessary. This will be supplemented with graphical posterior predictive checks (103, 104) focused on the expected values/probabilities for each outcome and Pareto-smoothed importance sampling leave-one-out cross-validation focused on the effective number of parameters (*p_loo*) compared to the actual number of parameters in the model (105-107). Results from the Bayesian analyses will be reported in accordance with the Reporting Of Bayes Used in clinical STudies (ROBUST) criteria (108).

12.8 Heterogenous treatment effects

Adaptations and primary analyses of EMPRESS will solely be based on average treatment effects calculated as described above. After the trial is stopped, we will also assess whether heterogenous treatment effects (HTE) for the primary outcome are present according to several pre-defined baseline characteristics (**Table 6).**

Table 6: Heterogeneity of the intervention effects on the primary outcome will be analysed based on baseline characteristics.

Baseline characteristics	Definition	Expected direction of interaction
Disease severity	Baseline disease severity according to the Simplified Mortality Score for the Intensive Care Unit (SMS-ICU; assessed on the continuous scale) (109)	Larger intervention effect in participants with higher disease severity
Primary site of infection	Participants with pulmonary versus gastrointestinal versus urinary tract versus skin or soft tissue versus other infections	Larger intervention effect in participants with urinary tract and skin or soft tissue infections compared to pulmonary, gastrointestinal, and other infections
Immunosuppression	Participants with known immunosuppression (i.e., haematological malignancy, metastatic cancer, recent or current use of immunosuppressive treatment, or transplanted organ) versus participants without immunosuppression	Larger intervention effect in participants with immunosuppression
Renal function	Highest plasma creatinine in the 24 hours prior to randomisation (assessed on the continuous scale)	Larger intervention effect in participants with worse baseline renal function (higher plasma creatinine)

Directions of interaction are expected to be in the same direction on both the absolute and relative scales, with the expected interactions being relative to the primary hypothesis of meropenem being superior.

HTE analyses of continuous baseline variables

The HTE analyses of continuous baseline variables (disease severity assessed by SMS-ICU and renal function assessed by highest plasma creatinine in the 24 hours prior to randomisation) will be conducted using logistic regression models similar to the primary analysis model, but also including main effect terms for the baseline variable of interest and the baseline variable of interest squared and interaction terms on the log-odds scale between the baseline variable of interest and the baseline variable of interest squared and the treatment effect. We expect that the distribution of plasma creatinine will be substantially right-skewed and will, thus, log2-transform the creatinine values prior to modelling. Both continuous baseline variables used in the HTE analyses will be mean-centred before being entered in the models (after log2-transformation of highest plasma creatinine) with visualization on the actual scale.

The resulting models will be used to make predictions of the expected probabilities in each treatment group across a grid of values for the baseline variable of interest (covering the full range of values) with *average* predictions made at each grid point by using the model fit with full dataset with the covariate of interest set to the grid value, the treatment indicator first set to piperacillin/tazobactam and then meropenem, and all other baseline covariates at their adjustment values. The posterior distributions of averages (means) of these predictions across all participants will be used to visualise the expected average probabilities of the outcome across the range of covariate values and further be used to calculate the expected average relative risks and risk differences across the same range, which will also be visualized. Median posterior values will be used for point estimates with 95% percentile-based credible intervals overlain along with visual presentations of the overall distributions of the baseline covariate of interest; the axis corresponding to the baseline covariate may be 'compressed' using an appropriate transformation to avoid giving undue space to rare values which may visually mislead. The results will be interpreted probabilistically across the range of values as a continuous measure of evidence.

HTE analyses of categorical baseline variables

The HTE analyses of categorical baseline variables (primary site of infection and immunosuppression) will be conducted using hierarchical logistic regression models similar with the primary models but including (uncorrelated) random effects for the intercept and treatment variable according to each category as previously done (110, 111). Hierarchical models partially pool data across subgroups and thus shrink effect estimates in each subgroup towards the overall estimate (112); the amount of shrinkage will be larger for

subgroups with more uncertain or extreme results (i.e., results further from the overall effect) and depends on the shrinkage prior used. A relatively weak shrinkage prior will primarily be used and given the overall expected size of the trial, we expect this to have limited influence on the results and primarily stabilise the models. The resulting models will be used to make predictions of the expected probabilities in each treatment group for each subgroup (each category of the baseline covariate of interest assessed); *average* predictions will be made in each subgroup by using the model fit with the full dataset, with the covariate of interest set to each subgroup in turn, the treatment indicator first set to piperacillin/tazobactam and then meropenem, and all other baseline covariates at their adjustment values. The posterior distributions of averages (means) of these predictions across all participants will be used to visualise the expected average probabilities of the outcome for each subgroup and further be used to calculate and visualise the expected average relative risks and risk differences for each subgroup. Median posterior values will be used for point estimates with 95% percentile-based credible intervals overlain along with visual presentations of the overall distributions of the baseline covariate of interest. The results will be interpreted probabilistically for each subgroup as a continuous measure of evidence.

12.9 Missing data handling

We will continuously monitor missing data in the electronic case report form (eCRF) and contact sites where missing data is prevalent to ensure as complete data as possible. We expect limited missing data for the primary outcome and most secondary outcomes (except the health-related quality-of-life outcomes, where moderate amounts of missing data may occur), and very few missing data for all covariates included in the primary analyses listed above. The proportions of missing data will be presented in the trial report. If there are no missing data for one or more variables included in an analysis, it will be conducted using complete cases only; otherwise, multiple imputation (MI) will be used regardless of the amount of missing data (113-115). We will assume that data are missing at random (MAR), as data being missing completely at random (MCAR) or fully not missing at random (NMAR) is unlikely in large trials with reasonable amounts of covariates and outcomes collected (116), and as MI increases power in both scenarios and generally reduces bias compared to complete case analysis in NMAR scenarios (116). We will generate 25 imputed datasets separately in each group using chained equations with the predictive mean matching and binary/ordinal/polytomous logistic regression methods, including all baseline variables listed in section 11.4 and **Table S1** (appendix 7, section 19.7) and all outcomes where data collection has concluded for all randomised participants. The continuous variables used in the HTE analyses (section 12.8) will be transformed (highest plasma creatinine only) and mean-centred before being entered in the imputation

models. Where MI is used, we will not present complete case analysis results, as these are at higher risk of being biased and thus adds limited value (116). If MI is used, all Bayesian models will be fit separately in each imputed dataset; models will be diagnosed separately in each dataset, while posteriors will be pooled and summarised with the required effective sample sizes applying to the pooled posteriors.

Where missing data are present, we will present *best-worst* and *worst-best* case scenarios as sensitivity analyses (115); for binary outcomes, we will impute the absence/presence of the outcome in both scenarios; for continuous outcomes with minimum/maximum limits, these will be used (with the lowest observed value used for missing EQ-5D-5L index values). Missing covariate data will be singly imputed using the median observed value for numerical variables and the most frequent category for categorical/binary variables in the *best-worst* and *worst-best* scenario analyses. Where data are only partially missing (e.g., for days alive without life support or days alive and out of hospital, where data may only be missing for some days, or for EQ-5D-5L, where data may only be missing for specific domains), the partially available data will be used. As such, we will truncate the imputed values for days alive without life support and days alive and out of hospital (in both analyses using MI and best-worst/worst-best case scenarios) to the possible range of values according to the partially observed data for each participant directly within the MI procedure. For EQ-5D-5L index values, these will be calculated following imputation of missing data for the EQ-5D-5L domains. For outcomes where non-survivors are assigned a special value (days alive without life support, days alive and out of hospital, and HRQoL outcomes), this will be done after the MI procedure to avoid overly large influence of the special values assigned to non-survivors on the imputation of values for survivors.

The adaptive analyses will be conducted using either complete case analysis or after MI according to the criteria and procedures outlined above, although no sensitivity analyses will be conducted for the adaptive analyses. MI models for the adaptive analyses will only include the primary outcome and all adjustment variables included in the primary analyses of the primary outcome (section 12.7), except if the last adaptive analysis includes the maximum sample size and thus coincides with the final analysis (section 12.4).

12.10 Overview of analyses including secondary and sensitivity analyses

This section summarises of the different analyses supplementing the primary, including secondary analyses, sensitivity analyses, and analyses of heterogeneous treatment effects (HTE) outlined in the rest of section 12. All secondary/sensitivity/HTE analyses will only be conducted for the primary outcome with all other analysis choices reflecting the primary analysis, except where otherwise specified.

- Adaptive analyses: primary outcome only, no sensitivity analyses

- Secondary analyses:
    - All outcomes: *per-protocol* analyses
    - All outcomes: analyses restricted to the population of participants receiving adequate empirical treatment (i.e., the intention-to-treat population excluding those with bacteria not fully sensitive to the allocated treatment in baseline cultures)
- Sensitivity analyses:
    - All outcomes: different priors (more informative, less informative, and if relevant also evidence-based priors)
    - Days alive without life support and days alive and out of hospital at day 30 and 90: no penalisation of death (actual values used for non-survivors)
    - Health-related quality of life (EQ-5D-5L index values and EQ VAS at day 180): analysis conducted in survivors only and analysis conducted for all participants (including non-survivors) using the Danish value set for all participants (117)
    - All outcomes: best-worst/worst-best case analyses (omitted if there are no missing data for an outcome)
- Analyses of heterogeneous treatment effects:
    - Primary outcome only, primary set of priors only.

12.11 Conclusions and statistical interpretation of the evidence

The primary conclusion of EMPRESS will be based on whether a stopping rule has been triggered, with stopping rules for superiority/inferiority calibrated to control the overall type 1 error rate for the primary outcome at 5% (section 12.4). Regardless of whether EMPRESS is stopped for superiority/inferiority, or practical equivalence, the results will be interpreted probabilistically, i.e., based on the probabilities of differences and different effect sizes, without any threshold for statistical significance. Secondary outcomes, secondary analyses, and analyses of heterogeneous treatment effects will similarly be interpreted probabilistically without statistical significance thresholds. As there are no thresholds for superiority (or statistical significance) or practical equivalence for any of the secondary outcomes or analyses, there will be no multiplicity corrections, but the results will be interpreted cautiously considering the number of outcomes and analyses.

13 Quality control and quality assurance

The sponsor and his delegates will be responsible for organising the trial sites, including education of the primary investigators. This education of primary investigators will be documented in the trial master file. After initiation of trial sites, the primary investigators will be responsible for all trial-related procedures at their sites, including education of staff in trial-related procedures, recruitment, follow-up of participants, screening in the eCRF, and data entry in the eCRF. Clinical staff at the trial sites will be responsible for screening of eligible patients and for the treatment of trial participants.

13.1 Monitoring

The trial will be externally monitored according to the GCP Directive and a monitoring and data verification plan. The monitoring and data verification plan will be developed by the sponsor and the GCP-unit of Copenhagen University Hospital and adhered to by the GCP staff monitoring all trial sites.
The GCP monitors and the relevant competent authorities (e.g., the Danish Medicines Agency) overseeing drug trials will have access to the participants electronic patient record for on-site monitoring of the trial, including monitoring of consent forms and source data verification. In addition, we will use central monitoring of sites through the eCRF, including monitoring of adherence to the protocol.

13.2 Drug traceability measures

The registration of the BATCH/LOT numbers, the expiry dates of the piperacillin/tazobactam, meropenem, and saline used, and the identity of the clinician administering the trial interventions will be registered as per standard practice at the sites (e.g., in the participant's electronic health records), i.e., if the BATCH/LOT numbers and the expiry date is not registered as per standard clinical practice, it will not be registered. Consequently, these data will not be registered in the eCRF but registered according to usual clinical practice and applicable local regulations. We believe that this is a safe procedure because both the piperacillin/tazobactam, meropenem, and saline used in the EMPRESS trial are 1) shelf-medication from the hospitals' pharmacies, 2) well-known drugs that have been in clinical use for many years, and 3) the safety of single doses cannot be questioned. The same procedure was approved by the Danish Medicines Agency in the COVID STEROID 2 (EudraCT no. 2020-003363-25) (118), COVID STEROID (EudraCT no. 2020-001395-15) (119) and CLASSIC trials (EudraCT no. 2018-000404-42) (36).

14 Legal and organisational aspects

14.1 Finance

Trial funding

The trial is funded by grants from the Independent Research Fund Denmark (4,220,996 DKK), the Research Council at Rigshospitalet (2,040,000 DKK), the Research Fund for Health Research of the Capital Region of Denmark (1,740,428 DKK), the Beckett Foundation (100,000 DKK), Læge Inger Goldmanns Fond (100,000 DKK), and Grosserer Jakob Ehrenreich og Hustru Grete Ehrenreichs Fond (512,000 DKK). The trial uses infrastructure and methodology developed as part of the Intensive Care Platform Trial (INCEPT) programme (www.incept.dk), funded by the Novo Nordisk Foundation (29,985,000 DKK) and Sygeforsikringen "danmark" (7,300,000 DKK) and supported by Grosserer Jakob Ehrenreich og Hustru Grete Ehrenreichs Fond (DKK 200,000), Dagmar Marshalls Fond (DKK 93,516), and Savværksejer Jeppe Juhl og hustru Ovita Juhls Mindelegat (DKK 1,000,000). The funding organisations have and will not be involved in the design, conduct, analyses, or reporting of the trial nor will it have ownership of the data.

Compensation

The trial sites will receive compensation (case money) per included participant to cover the expenses of participant enrolment. This compensation will be based on estimated time consumption and available funding and may be adjusted during the course of the trial. Case money may differ between countries and/or regions to account for variations in the overall workload related to participant enrolment. For instance, we plan to automate some data entry processes, reducing the overall workload, and this will be implemented differentially across sites, regions, and countries according to what is technically possible. The compensation per participant may be adjusted accordingly.

Insurance

In Denmark, the trial participants are covered by the Danish Law 'Lov om Patientskadeerstatning'. Insurance in other participating countries will be specified in the respective local protocols.

Support

Simulations were performed on the UCloud (docs.cloud.sdu.dk) interactive high-performance computing system managed by the eScience Center at the University of Southern Denmark, and thus partially supported by the Danish e-infrastructure Consortium (DeiC) National HPC (DeiC-KU-S1-000114).

15 Additional details

15.1 Protocol modifications

Protocol modifications will be approved by the competent authorities before implementation. If necessary, the information to trial participants will be changed and approved accordingly. Upon approval, notifications about relevant protocol modifications will be sent to all primary investigators and monitors.

15.2 Trial results and reporting

All trial results of the clinical outcomes, whether positive, negative, neutral, or inconclusive, will be published in a peer-reviewed medical journal. The results of the feasibility phase will be published in a separate publication also in a peer-reviewed medical journal unless the trial is deemed infeasible; in this case, we will report both clinical and feasibility outcomes in the same publication. Furthermore, the results will be published via CTIS, at ClinicalTrials.gov, and on the trial website. For the reporting, we will adhere to the CONSORT-ACE statement (31) and the ROBUST criteria for the reporting of the Bayesian analyses (108). The management committee will write the trial report(s).

Authorship will be granted according to the guidelines from the International Committee of Medical Journal Editors (ICMJE) (120). The listing of authors will be granted according to the overall contribution to the trial design and conduct. The management committee may grant additional authorships depending on personal input as per the Vancouver criteria (120). Investigators on sites may also be granted authorship on sub-study publications if they contribute significantly as per the Vancouver criteria (120).

The IDMSC and investigators not qualifying for authorship will be acknowledged with their names in an appendix to the final manuscript. The funding sources will also be acknowledged, but they will have no influence on the data handling or analyses, the writing of the manuscript, or the decision to publish.

15.3 Secondary studies

Secondary studies will be encouraged if they do not hamper the completion of the main protocol. Specific protocols for any sub-studies will be submitted to and approved by the competent authorities (if needed) and the management committee before the commencement of such studies.

15.4 Intellectual property rights

The sponsor site owns the trial data.

15.5 Data management and data sharing

Data management

All clinical trial data will be collected systematically from electronic patient records (EPRs) using a dedicated electronic case report form (eCRF) with pre-defined variables. Data will be collected both prospectively and retrospectively with minimal delay to minimise the potential for errors and omissions. We are developing the data-scientific infrastructure for partial automatic data capture from, e.g., EPRs and expect to start using this for data collection at relevant sites in the course of this trial, at which point the protocol will be amended to fully document the logic and codification used.

To maintain data accuracy and integrity, regular checks will be conducted throughout the duration of the trial. Inconsistencies, missing data, and outliers will be identified and handled as appropriate. All collected data will be stored in a secure and centralised electronic database with regular backups during the course of the trial, to prevent unauthorised access. Access control, activity logging and soft deletion will be implemented to document changes in the database and enable reversal of accidental or otherwise erroneous data entry.

Upon trial completion, the final data will be processed and saved in a format appropriate for persistent storage, archived an appropriate location (e.g., a secure, logged network drive in Region Hovedstaden) for 25 years. The trial sponsor and their delegates will have access to the final trial dataset, whereas investigators will have access to data for participants enrolled at their respective sites.

Data sharing

An anonymised version of the final dataset (without personal, identifiable information, with timestamps replaced by relative time differences with respect to the time of randomisation, and other measures as deemed relevant) may be shared with other researchers following a reasonable request (i.e., a research proposal outlining the objectives, methodologies, and plans for data usage) and subsequent approval by the management committee. Any sharing of data that is not considered anonymised will be after the necessary approvals; alternatively, aggregation, scrambling, or synthetic datasets (121) (i.e., datasets with similar structure and attempts to preserve the overall relationships between variables as the original dataset) may be shared. Data will generally only be shared after a grace period of at least 9 months following initial publication of results based on the data. Approved researchers will sign appropriate

agreements to ensure compliance with the approved purpose and ethical and eventual legal requirements. Participants will be informed about the possibility of data sharing during the informed consent process.

Code sharing

Analysis code may be shared with other researcher after reasonable request and approval by the management committee.

15.6 Organisational framework

The EMPRESS trial will be conducted and managed by the sponsor, the management committee (p. 1), the dedicated trial staff, the investigators, and the Research Unit at the Department of Intensive Care, Rigshospitalet.

15.7 Conflicts of interest

The financial and other competing interests will be stated for all co-authors in each trial publication.

Conflicts of interest for the management committee involved in writing of the protocol

The Department of Intensive Care at Rigshospitalet has received grants from the Novo Nordisk Foundation, Pfizer, and Sygeforsikringen “danmark”.

Anders Perner has participated in advisory boards for Novartis.

Marie Helleberg has participated in advisory boards for GlaxoSmithKline, AstraZeneca, Pfizer, Sobi, Roche, and MSD.

16 Estimated timeline

- Medio 2024: authority approvals and first participant randomised
- Medio 2025: feasibility phase analysis concluded
- Ultimo 2028: expected inclusion of the last participant if trial continues to the expected sample size in the small-difference scenario (i.e., the largest expected sample size under the three different scenarios assessed) (section 19.9)
- Ultimo 2032: expected inclusion of the last participant if the trial continues to the maximum sample size (n=14,000) (section 19.9)
- Approximately 3 months after inclusion of the last participant: primary report on 30-day outcomes submitted
- Approximately 6 months after inclusion of the last participant: report on 90-day outcomes submitted
- Approximately 9 months after inclusion of the last participant: report on 180-day outcomes submitted

17 Summary of changes

This section summarises all changes to the protocol after initial submission for approval.

Version 1.10, 2024.03.22:

- Clarifications with regards to the enrolment and informed consent procedures being according to the applicable legislation for clinical trials conducted in emergency situations in multiple places (section 4.4; section 8.2; section 11.2; and appendix 6, section 19.5, which has been renamed). Further, the Danish procedure for enrolment and informed consent has been removed from the main protocol, as per request from the competent authorities. Procedures for enrolment and informed consent in each participating country will be submitted as separate documents to the competent authorities and made available on the trial website prior to start of enrolment in each participating country.
- Clarifications that de-escalation to another antibiotic than the two assessed in the trial due to clinical improvement (including in situations where no bacteria have been found in microbiological cultures) is permitted, will be registered, but is not considered a protocol violation (section 4.3; section 6.5; section 8.3; section 8.4; section 8.5; section 8.9; section 11.4; and appendix 4, section 19.4).
- Additional clarification in section 6.5 that HRQoL data will only be obtained through proxies in situations where participants cannot respond themselves.
- Corrected an error in the mock inclusion flowchart (the number of participants analysed was erroneously not included in one arm) (appendix 3, section 19.3).
- Added references to two studies previously referred to as 'unpublished data from our group'.
- Minor corrections and semantic edits in multiple places.

Version 1.9, 2024.02.28:

- Added ClinicalTrials.gov registration number to the first page.
- The estimated timeline in section 1 and section 16 has been updated.
- Added reference to guidelines from the Danish Society of Infectious Diseases with regards to meropenem doses in section 4.3.
- Added additional justification for the expected and maximum sample sizes in the new section 6.1 along with additional justification for the size of the integrated feasibility phase in section 6.3 and section 9.2.

- Clarification regarding collection of HRQoL data from participants' next of kin where relevant in section 6.5.
- Clarification that dose adjustments are allowed at the treating clinicians discretion, including but not limited to the provided examples in sections 8.4 and 8.5.
- Clarifications on deviations from the protocol by the clinical team in section 8.8.
- Corrected definition of SARs in section 10.1 and details on classification of expected SARs and SUSARs in section 10.3, correction on the reporting of SUSARs (reporting via the EudraVigilance system).
- Clarification that none of the drugs assessed are subject to stricter reporting requirements in Denmark in section 10.2.
- Added that investigators will be able to report any SAEs immediately at their own discretion (section 10.2) and that such events will be included in the annual safety reports (section 10.3).
- Clarification that assessments of causality between the intervention and SUSARs will be made, and that all *expected* SARs will be considered causal in section 10.3.
- A dedicated section providing a summary of the risk-benefit assessment has been added as section 10.5.
- Clarifications to the screening procedure in section 11.1 with regards to the role of the trained trial investigators and the treating clinical doctor.
- Added chronic use of renal replacement therapy to the baseline variables collected (section 11.4; appendix 4, section 19.4; appendix 7, section 19.7).
- Additional justification for the clinical condition constituting an emergency situation and why enrolment before informed consent is required has been provided in section 11.2 and appendix 6, section 19.6.
- Additional details on the handling of partially missing data have been added to section 12.9.
- Updated conflicts of interest (section 15.7).
- Minor corrections and semantic edits in multiple places not leading to any changes in meaning.

Version 1.8, 2023.12.19: first version submitted for approval.

18 References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):801-10.
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):1181-247.
3. Yealy DM, Mohr NM, Shapiro NI, Venkatesh A, Jones AE, Self WH. Early Care of Adults With Suspected Sepsis in the Emergency Department and Out-of-Hospital Environment: A Consensus-Based Task Force Report. Ann Emerg Med. 2021;78(1):1-19.
4. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. Lancet. 2020;395(10219):200-11.
5. GBD Antimicrobial Resistance Collaborators. Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2023;400(10369):2221-48.
6. Venkatesh B, Finfer S, Cohen J, Rajbhandari D, Arabi Y, Bellomo R, et al. Adjunctive Glucocorticoid Therapy in Patients with Septic Shock. N Engl J Med. 2018;378(9):797-808.
7. De Bus L, Depuydt P, Steen J, Dhaese S, De Smet K, Tabah A, et al. Antimicrobial de-escalation in the critically ill patient and assessment of clinical cure: the DIANA study. Intensive Care Med. 2020;46(7):1404-17.
8. Meier N, Munch MW, Granholm A, Perner A, Hertz FB, Venkatesh B, et al. Empirical carbapenems or piperacillin/tazobactam for infections in intensive care – an international retrospective cohort study. Acta Anaesthesiologica Scandinavica. 2024; accepted, awaiting publication.
9. Eljaaly K, Enani MA, Al-Tawfiq JA. Impact of carbapenem versus non-carbapenem treatment on the rates of superinfection: A meta-analysis of randomized controlled trials. J Infect Chemother. 2018;24(11):915-20.
10. McLaughlin M, Advincula MR, Malczynski M, Qi C, Bolon M, Scheetz MH. Correlations of antibiotic use and carbapenem resistance in enterobacteriaceae. Antimicrob Agents Chemother. 2013;57(10):5131-3.
11. Harris PNA, Tambyah PA, Lye DC, Mo Y, Lee TH, Yilmaz M, et al. Effect of Piperacillin-Tazobactam vs Meropenem on 30-Day Mortality for Patients With E coli or Klebsiella pneumoniae Bloodstream Infection and Ceftriaxone Resistance: A Randomized Clinical Trial. JAMA. 2018;320(10):984-94.
12. Henderson A, Paterson DL, Chatfield MD, Tambyah PA, Lye DC, De PP, et al. Association Between Minimum Inhibitory Concentration, Beta-lactamase Genes and Mortality for Patients Treated With Piperacillin/Tazobactam or Meropenem From the MERINO Study. Clin Infect Dis. 2021;73(11):e3842-e50.
13. Son SK, Lee NR, Ko JH, Choi JK, Moon SY, Joo EJ, et al. Clinical effectiveness of carbapenems versus alternative antibiotics for treating ESBL-producing Enterobacteriaceae bacteraemia: a systematic review and meta-analysis. J Antimicrob Chemother. 2018;73(10):2631-42.
14. Sfeir M, Askin G, Christos P. Beta-lactam/ beta-lactamase inhibitors versus carbapenem for bloodstream infections due to extended spectrum beta-lactamase producing Enterobacteriaceae: Systematic review and meta-analysis. International journal of antimicrobial agents. 2018;52(5):554-70.
15. Muhammed M, Flokas ME, Detsis M, Alevizakos M, Mylonakis E. Comparison Between Carbapenems and beta-Lactam/beta-Lactamase Inhibitors in the Treatment for Bloodstream Infections Caused by Extended-Spectrum beta-Lactamase-Producing Enterobacteriaceae: A Systematic Review and Meta-Analysis. Open Forum Infect Dis. 2017;4(2):ofx099.

16. Vardakas KZ, Tansarli GS, Rafailidis PI, Falagas ME. Carbapenems versus alternative antibiotics for the treatment of bacteraemia due to Enterobacteriaceae producing extended-spectrum beta-lactamases: a systematic review and meta-analysis. J Antimicrob Chemother. 2012;67(12):2793-803.
17. Harris PN, Wei JY, Shen AW, Abdile AA, Paynter S, Huxley RR, et al. Carbapenems versus alternative antibiotics for the treatment of bloodstream infections caused by Enterobacter, Citrobacter or Serratia species: a systematic review with meta-analysis. J Antimicrob Chemother. 2016;71(2):296-306.
18. Cheng MP, Lee RS, Cheng AP, De L'etoile-Morel S, Demir K, Yansouni CP, et al. Beta-Lactam/Beta-Lactamase Inhibitor Therapy for Potential AmpC-Producing Organisms: A Systematic Review and Meta-Analysis. Open Forum Infect Dis. 2019;6(7).
19. An MM, Zou Z, Shen H, Zhang JD, Chen ML, Liu P, et al. Ertapenem versus piperacillin/tazobactam for the treatment of complicated infections: a meta-analysis of randomized controlled trials. BMC Infect Dis. 2009;9:193.
20. Shiber S, Yahav D, Avni T, Leibovici L, Paul M. beta-Lactam/beta-lactamase inhibitors versus carbapenems for the treatment of sepsis: systematic review and meta-analysis of randomized controlled trials. J Antimicrob Chemother. 2015;70(1):41-7.
21. Edwards SJ, Clarke MJ, Wordsworth S, Emmas CE. Carbapenems versus other beta-lactams in treating severe infections in intensive care: a systematic review of randomised controlled trials. Eur J Clin Microbiol Infect Dis. 2008;27(7):531-43.
22. Munch MW, Granholm A, Jonsson AB, Sjövall F, Helleberg M, Hertz FB, et al. Piperacillin/tazobactam versus Carbapenems in Patients with Severe Bacterial Infections: A Systematic Review with Meta-analysis. Acta Anaesthesiologica Scandinavica.n/a(n/a).
23. Dosages used to define breakpoints. EUCAST Clinical Breakpoint Tables v. 13.0, valid from 2023-01-01. Available from: https://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/Dosages_v_13.0_Breakpoint_Tables.pdf. Accessed on January 17 2023.
24. Dansk selskab for infektionsmedicin. Sepsis guidelines 2021. Available from: https://www.infmed.dk/guidelines#sepsis_guidelines_(2021).pdf. Accessed on 25 February 2024.
25. Granholm A, Kaas-Hansen BS, Lange T, Schjorring OL, Andersen LW, Perner A, et al. An overview of methodological considerations regarding adaptive stopping, arm dropping, and randomization in clinical trials. J Clin Epidemiol. 2022;153:45-54.
26. Den Europæiske Unions Tidende. Europa-parlamentets og rådets forordning (EU) Nr. 536/2014 af 16. april 2014 om kliniske forsøg med humanmedicinske lægemidler og om ophævelse af direktiv 2001/20/EF (EØS-relevant tekst). Version 27 May 2014. Available from: https://eur-lex.europa.eu/legal-content/DA/TXT/PDF/?uri=CELEX:32014R0536&from=EN;%20%E2%80%9DREGULATION%20(EU)%20No%20536/2014%20OF%20THE%20EUROPEAN%20PARLIAMENT%20AND%20OF%20THE%20COUNCIL%20of%2016%20April%202014%20on%20clinical%20trials%20on%20medicinal%20products%20for%20human%20use,%20and%20repealing%20Directive%202001/20/EC%E2%80%9D. Accessed on 25 February 2024.
27. Indenrigs- og Sundhedsministeriet. Bekendtgørelse om kliniske forsøg med lægemidler. BEK nr 12 af 06/01/2022. Available from: https://www.retsinformation.dk/eli/lta/2022/12. Accessed on 25 February 2024.
28. World Medical Association. WMA Declaration of Helsinki - Ethical Principles for Medical Research involving Human Subjects. Available from: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/20132013/. Accessed on 5 July 2023. 2013.
29. European Medicines Agency. Science Medicines Health. Good Clinical Practice. Available from: https://www.ema.europa.eu/en/human-regulatory/research-development/compliance/good-clinical-practice#international-clinical-trials-section. Accessed on March 18 2020.
30. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gotzsche PC, Krleza-Jeric K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med. 2013;158(3):200-7.

31. Dimairo M, Pallmann P, Wason J, Todd S, Jaki T, Julious SA, et al. The adaptive designs CONSORT extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. Trials. 2020;21(1):528.
32. European Medicines Agency. FAQs. How to manage a clinical trial. CTIS Training Programme - Module 05. Version 1.3, March 2022. Available from: https://www.ema.europa.eu/en/documents/other/faqs-how-manage-ct-ctis-training-programme-module-05_en.pdf. Accessed on 30 October 2023.
33. Krag M, Marker S, Perner A, Wetterslev J, Wise MP, Schefold JC, et al. Pantoprazole in Patients at Risk for Gastrointestinal Bleeding in the ICU. N Engl J Med. 2018;379(23):2199-208.
34. Schjørring OL, Klitgaard TL, Perner A, Wetterslev J, Lange T, Siegemund M, et al. Lower or Higher Oxygenation Targets for Acute Hypoxemic Respiratory Failure. New England Journal of Medicine. 2021;384(14):1301-11.
35. COVID STEROID 2 Trial Group, Munch MW, Myatra SN, Vijayaraghavan BKT, Saseedharan S, Benfield T, et al. Effect of 12 mg vs 6 mg of Dexamethasone on the Number of Days Alive Without Life Support in Adults With COVID-19 and Severe Hypoxemia: The COVID STEROID 2 Randomized Trial. JAMA. 2021.
36. Meyhoff TS, Hjortrup PB, Wetterslev J, Sivapalan P, Laake JH, Cronhjort M, et al. Restriction of Intravenous Fluid in ICU Patients with Septic Shock. N Engl J Med. 2022;386(26):2459-70.
37. Andersen-Ranberg NC, Poulsen LM, Perner A, Wetterslev J, Estrup S, Hästbacka J, et al. Haloperidol for the Treatment of Delirium in ICU Patients. New England Journal of Medicine. 2022;387(26):2425-35.
38. Young PJ, Arabi M, Bagshaw SM, Bellomo R, Fujii T, Haniffa R, et al. Protocol and statistical analysis plan for the mega randomised registry trial research program comparing conservative versus liberal oxygenation targets in adults receiving unplanned invasive mechanical ventilation in the ICU (Mega-ROX). Crit Care Resusc. 2022;24(2):137-49.
39. Altman DG, Bland JM. How to randomise. BMJ. 1999;319(7211):703-4.
40. National Center for Complementary and Integrative Health. Pilot Studies: Common Uses and Misuses. Available from: https://www.nccih.nih.gov/grants/pilot-studies-common-uses-and-misuses. Accessed on 25 February 2024.
41. Totton N, Lin J, Julious S, Chowdhury M, Brand A. A review of sample sizes for UK pilot and feasibility studies on the ISRCTN registry from 2013 to 2020. Pilot and Feasibility Studies. 2023;9(1):188.
42. Teresi JA, Yu X, Stewart AL, Hays RD. Guidelines for Designing and Evaluating Feasibility Pilot Studies. Medical Care. 2022;60(1):95-103.
43. Danish Medicines Agency. Summary of Product Characteristics for Piperacillin/Tazobactam "Reig Jofre". February 8 2021. Available from: https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fspcweb.produktresume.dk%2FSPCREPL%2FHuman%2FP%2FPiperacillin-Tazobactam%2520Reig%2520Jofre%2C%2520pulver%2520til%2520infusionsv%25c3%25a6ske%2C%2520opl%25c3%25b8sning%25202%2520g%252b0%2C25%2520g%2520og%25204%2520g%252b0%2C5%2520g.doc&wdOrigin=BROWSELINK. Accessed on February 1 2023.
44. Danish Medicines Agency. Summary of Product Characteristics for Meropenem "Fresenius Kabi". August 1 2022. Available from: https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fspcweb.produktresume.dk%2FSPCREPL%2FHuman%2FM%2FMeropenem%2520Fresenius%2520Kabi%2C%2520pulver%2520til%2520injektions-%2520og%2520infusionsv%25c3%25a6ske%2C%2520opl%25c3%25b8sning%2520500%2520mg%2520og%25201000%2520mg.doc&wdOrigin=BROWSELINK. Accessed on February 1 2023.
45. Anthon CT, Granholm A, Perner A, Laake JH, Moller MH. No firm evidence that lack of blinding affects estimates of mortality in randomized clinical trials of intensive care interventions: a systematic review and meta-analysis. J Clin Epidemiol. 2018;100:71-81.

46. Moustgaard H, Clayton GL, Jones HE, Boutron I, Jorgensen L, Laursen DRT, et al. Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. BMJ. 2020;368:l6802.
47. Martin GL, Trioux T, Gaudry S, Tubach F, Hajage D, Dechartres A. Association Between Lack of Blinding and Mortality Results in Critical Care Randomized Controlled Trials: A Meta-Epidemiological Study. Crit Care Med. 2021;49(10):1800-11.
48. Cheng LJ, Engel L, Chen LA, Soh SZY, Koh GC-H, Luo N. Using EQ-5D for Proxy Assessment of Health-Related Quality of Life in Residential Care Facilities: A Systematic Review of Feasibility and Psychometric Properties. Journal of the American Medical Directors Association.
49. Pickard AS, Knight SJ. Proxy Evaluation of Health-Related Quality of Life: A Conceptual Framework for Understanding Multiple Proxy Perspectives. Medical Care. 2005;43(5):493-9.
50. Granholm A, Kjaer MN, Munch MW, Myatra SN, Vijayaraghavan BKT, Cronhjort M, et al. Long-term outcomes of dexamethasone 12 mg versus 6 mg in patients with COVID-19 and severe hypoxaemia. Intensive Care Med. 2022;48(5):580-9.
51. Mortensen CB, Andersen-Ranberg NC, Poulsen LM, Granholm A, Rasmussen BS, Kjær M-BN, et al. Long-term outcomes with haloperidol versus placebo in acutely admitted adult ICU patients with delirium. Intensive Care Medicine. 2024;50(1):103-13.
52. Kjær M-BN, Meyhoff TS, Sivapalan P, Granholm A, Hjortrup PB, Madsen MB, et al. Long-term effects of restriction of intravenous fluid in adult ICU patients with septic shock. Intensive Care Medicine. 2023;49(7):820-30.
53. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med. 1996;22(7):707-10.
54. Sjovall F, Alobaid AS, Wallis SC, Perner A, Lipman J, Roberts JA. Maximally effective dosing regimens of meropenem in patients with septic shock. J Antimicrob Chemother. 2018;73(1):191-8.
55. Antibiotikavejledning, Rigshospitalets intensivafdelinger. Available from: https://vip.regionh.dk/VIP/Admin/vipportal.nsf/index.html. Accessed on 4 April 2023.
56. Gilbert DN, Chambers HF, Saag MS, Pavia AT, Boucher HW, Black D, et al. The Sanford Guide to Antimicrobial Therapy. 53rd edition. 2023.
57. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727-36.
58. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Integrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6(R2). Current Step 4 version, November 9 2016.
59. The Danish Medicines Agency. Lægemiddelstyrelsens vejledning om risikotilpasset bivirkningsregistrering og rapportering i kliniske forsøg med lægemidler under forordning (EU) nr. 536/2014. Version 1.0, October 2023. Available from: https://laegemiddelstyrelsen.dk/da/nyheder/2023/risikotilpasset-bivirkningsregistrering-og-rapportering-i-kliniske-forsoeg-med-laegemidler-er-nu-muligt/. Accessed on 16 November 2023.
60. Risk proportionate approaches in clinical trials. Recommendations of the expert group on clinical trials for the implementation of Regulation (EU) No 536/2014 on clinical trials on medicinal products for human use. 25 April 2017. Available from: https://health.ec.europa.eu/medicinal-products/eudralex/eudralex-volume-10_en. Accessed on 23 November 2023.
61. Danish Medicines Agency. Medicines with stricter reporting requirements for doctors, dentists, veterinarians, midwives and prescribing pharmacists. 20 December 2023. Available from: https://laegemiddelstyrelsen.dk/en/sideeffects/side-effects-of-medicines/medicines-with-stricter-reporting-requirements/. Accessed on 25 February 2024.

62. 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. Critical Care Medicine. 2021;49(11):e1063-e143.
63. Rockwood K, Song X, MacKnight C, Bergman H, Hogan DB, McDowell I, et al. A global clinical measure of fitness and frailty in elderly people. CMAJ. 2005;173(5):489-95.
64. Rockwood K, Theou O. Using the Clinical Frailty Scale in Allocating Scarce Health Care Resources. Canadian Geriatrics Journal. 2020;23(3):254-9.
65. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline. Statistical Principles for Clinical Trials E9. Current Step 4 version, February 5 1988. Available from: https://database.ich.org/sites/default/files/E9_Guideline.pdf. Accessed on February 1 2023.
66. European Medicines Agency. Complex clinical trials – Questions and answers. Version 23 May 2022. Available from: https://health.ec.europa.eu/system/files/2022-06/medicinal_qa_complex_clinical-trials_en.pdf. Accessed on 25 January 2023.
67. U.S. Department of Health and Human Services Food and Drug Administration. Adaptive Designs for Clinical Trials of Drugs and Biologics. Guidance for Industry. November 2019. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry. Accessed on 25 January 2023.
68. Talisa VB, Yende S, Seymour CW, Angus DC. Arguing for Adaptive Clinical Trials in Sepsis. Front Immunol. 2018;9:1502.
69. Granholm A, Alhazzani W, Derde LPG, Angus DC, Zampieri FG, Hammond NE, et al. Randomised clinical trials in critical care: past, present and future. Intensive Care Med. 2022;48(2):164-78.
70. Adaptive Platform Trials C. Adaptive platform trials: definition, design, conduct and reporting considerations. Nat Rev Drug Discov. 2019;18(10):797-807.
71. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31(4):337-50.
72. Hemming K, Javid I, Taljaard M. A review of high impact journals found that misinterpretation of non-statistically significant results from randomized trials was common. J Clin Epidemiol. 2022;145:112-20.
73. Lytsy P, Hartman M, Pingel R. Misinterpretations of P-values and statistical tests persists among researchers and professionals working with statistics and epidemiology. Ups J Med Sci. 2022;127.
74. Abrams D, Montesi SB, Moore SKL, Manson DK, Klipper KM, Case MA, et al. Powering Bias and Clinically Important Treatment Effects in Randomized Trials of Critical Illness. Crit Care Med. 2020;48(12):1710-9.
75. Granholm A, Lange T, Harhay MO, Perner A, Møller MH, Kaas-Hansen BS. Effects of sceptical priors on the performance of adaptive clinical trials with binary outcomes. Pharm Stat. 2024; accepted, awaiting publication.
76. Ridgeon EE, Bellomo R, Aberegg SK, Sweeney RM, Varughese RS, Landoni G, et al. Effect sizes in ongoing randomized controlled critical care trials. Crit Care. 2017;21(1):132.
77. Harhay MO, Wagner J, Ratcliffe SJ, Bronheim RS, Gopal A, Green S, et al. Outcomes and statistical power in adult critical care randomized trials. Am J Respir Crit Care Med. 2014;189(12):1469-78.
78. Cuthbertson BH, Scales DC. "Paying the Piper": The Downstream Implications of Manipulating Sample Size Assumptions for Critical Care Randomized Control Trials. Crit Care Med. 2020;48(12):1885-6.
79. Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ. 1995;311(7003):485.
80. Granholm A, Jensen AKG, Lange T, Kaas-Hansen BS. adaptr: an R package for simulating and comparing adaptive clinical trials. The Journal of Open Source Software. 2022;7(72):4284.
81. Wassmer G, Brannath B. Group Sequential and Confirmatory Adaptive Designs in Clinical Trials: Springer International Publishing; 2016.

82. Wang H, Rosner GL, Goodman SN. Quantifying over-estimation in early stopped clinical trials and the "freezing effect" on subsequent research. Clin Trials. 2016;13(6):621-31.
83. Liu S, Garrison SR. Overestimation of benefit when clinical trials stop early: a simulation study. Trials. 2022;23(1):747.
84. Walter SD, Guyatt GH, Bassler D, Briel M, Ramsay T, Han HD. Randomised trials with provision for early stopping for benefit (or harm): The impact on the estimated treatment effect. Stat Med. 2019;38(14):2524-43.
85. Granholm A, Lange T, Harhay MO, Jensen AKG, Perner A, Moller MH, et al. Effects of duration of follow-up and lag in data collection on the performance of adaptive clinical trials. Pharm Stat. 2023.
86. Viele K, Broglio K, McGlothlin A, Saville BR. Comparison of methods for control allocation in multiple arm studies using response adaptive randomization. Clin Trials. 2020;17(1):52-60.
87. Stan Development Team. Prior Choice Recommendations. Available from: https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations. Accessed on 4 July 2023.
88. Granholm A, Munch MW, Moller MH, Lange T, Perner A. Choice of priors: how much scepticism is appropriate? Intensive Care Med. 2022;48(3):372-3.
89. Chatton A, Le Borgne F, Leyrat C, Gillaizeau F, Rousseau C, Barbin L, et al. G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study. Sci Rep. 2020;10(1):9219.
90. Granholm A, Munch MW, Myatra SN, Vijayaraghavan BKT, Cronhjort M, Wahlin RR, et al. Dexamethasone 12 mg versus 6 mg for patients with COVID-19 and severe hypoxaemia: a pre-planned, secondary Bayesian analysis of the COVID STEROID 2 trial. Intensive Care Med. 2022;48(1):45-55.
91. U.S. Department of Health and Human Services Food and Drug Administration. Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products Guidance for Industry. May 2023. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized-clinical-trials-drugs-and-biological-products. Accessed on January 25 2023.
92. Granholm A, Kaas-Hansen BS, Lange T, Munch MW, Harhay MO, Zampieri FG, et al. Use of days alive without life support and similar count outcomes in randomised clinical trials – an overview and comparison of methodological choices and analysis methods. BMC Medical Research Methodology. 2023;23(1):139.
93. Schoenfeld DA, Bernard GR, Network ftA. Statistical evaluation of ventilator-free days as an efficacy measure in clinical trials of treatments for acute respiratory distress syndrome. Critical Care Medicine. 2002;30(8):1772-7.
94. Yehya N, Harhay MO, Curley MAQ, Schoenfeld DA, Reeder RW. Reappraisal of Ventilator-Free Days in Critical Care Research. Am J Respir Crit Care Med. 2019;200(7):828-36.
95. Contentin L, Ehrmann S, Giraudeau B. Heterogeneity in the definition of mechanical ventilation duration and ventilator-free days. Am J Respir Crit Care Med. 2014;189(8):998-1002.
96. Devlin N, Parkin D, Janssen B. Methods for Analysing and Reporting EQ-5D Data. Cham (CH)2020.
97. Jensen CE, Sorensen SS, Gudex C, Jensen MB, Pedersen KM, Ehlers LH. The Danish EQ-5D-5L Value Set: A Hybrid Model Using cTTO and DCE Data. Appl Health Econ Health Policy. 2021;19(4):579-91.
98. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. J Stat Softw. 2017;76.
99. Bürkner P-C. brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software. 2017;80(1):1 - 28.
100. Granholm A, Munch MW, Myatra SN, Vijayaraghavan BKT, Cronhjort M, Wahlin RR, et al. Higher vs Lower Doses of Dexamethasone in Patients with COVID-19 and Severe Hypoxia (COVID STEROID 2) trial: Protocol for a secondary Bayesian analysis. Acta Anaesthesiol Scand. 2021.

101. Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C. Rank-Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC (with Discussion). Bayesian Analysis. 2021;16(2):667-718, 52.
102. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. 3rd ed. CRC Press; 2014.
103. Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. Visualization in Bayesian workflow. arXiv preprint arXiv:170901449. 2017.
104. Kruschke JK. Doing Bayesian Data Analysis. 2nd ed. Academic Press; 2015.
105. Vehtari, V. Cross-validation FAQ. First version 2020-03-11. Last modified 2022-07-30. Available from: https://avehtari.github.io/modelselection/CV-FAQ.html. Accessed on February 1 2023.
106. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing. 2017;27(5):1413-32.
107. Lambert B. A Student's Gudie to Bayesian Statistics. 1st ed. SAGE Publications Ltd.; 2018.
108. Sung L, Hayden J, Greenberg ML, Koren G, Feldman BM, Tomlinson GA. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. J Clin Epidemiol. 2005;58(3):261-8.
109. Granholm A, Perner A, Krag M, Hjortrup PB, Haase N, Holst LB, et al. Development and internal validation of the Simplified Mortality Score for the Intensive Care Unit (SMS-ICU). Acta Anaesthesiologica Scandinavica. 2018;62(3):336-46.
110. Klitgaard TL, Schjørring OL, Lange T, Møller MH, Perner A, Rasmussen BS, et al. Lower versus higher oxygenation targets in critically ill patients with severe hypoxaemia: secondary Bayesian analysis to explore heterogeneous treatment effects in the Handling Oxygenation Targets in the Intensive Care Unit (HOT-ICU) trial. Br J Anaesth. 2022;128(1):55-64.
111. Granholm A, Marker S, Krag M, Zampieri FG, Thorsen-Meyer H-C, Kaas-Hansen BS, et al. Heterogeneity of treatment effect of prophylactic pantoprazole in adult ICU patients: a post hoc analysis of the SUP-ICU trial. Intensive Care Medicine. 2020;46(4):717-26.
112. Granholm A, Marker S, Krag M, Zampieri FG, Thorsen-Meyer HC, Kaas-Hansen BS, et al. Heterogeneity of treatment effect of stress ulcer prophylaxis in ICU patients: A secondary analysis protocol. Acta Anaesthesiol Scand. 2019;63(9):1251-6.
113. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011;45 (3).
114. Vesin A, Azoulay E, Ruckly S, Vignoud L, Rusinova K, Benoit D, et al. Reporting and handling missing values in clinical studies in intensive care units. Intensive Care Med. 2013;39(8):1396-404.
115. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol. 2017;17(1):162.
116. van Ginkel JR, Linting M, Rippe RCA, van der Voort A. Rebutting Existing Misconceptions About Multiple Imputation as a Method for Handling Missing Data. J Pers Assess. 2020;102(3):297-308.
117. Jensen CE, Sørensen SS, Gudex C, Jensen MB, Pedersen KM, Ehlers LH. The Danish EQ-5D-5L Value Set: A Hybrid Model Using cTTO and DCE Data. Applied Health Economics and Health Policy. 2021;19(4):579-91.
118. Munch MW, Myatra SN, Vijayaraghavan BKT, Saseedharan S, Benfield T, Wahlin RR, et al. Effect of 12 mg vs 6 mg of Dexamethasone on the Number of Days Alive Without Life Support in Adults With COVID-19 and Severe Hypoxemia: The COVID STEROID 2 Randomized Trial. Jama. 2021;326(18):1807-17.
119. Munch MW, Meyhoff TS, Helleberg M, Kjaer MN, Granholm A, Hjortso CJS, et al. Low-dose hydrocortisone in patients with COVID-19 and severe hypoxia: the COVID STEROID randomised, placebo-controlled trial. Acta Anaesthesiol Scand. 2021.

120. International Committee of Medical Journal Editors. Defining the Role of Authors and Contributors. Available from: https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html. Accessed on January 24 2023.
121. Nowok B, Raab GM, Dibben C. synthpop: Bespoke Creation of Synthetic Data in R. Journal of Statistical Software. 2016;74(11):1 - 26.
122. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.
123. Schoenfeld DA, Bernard GR, Network A. Statistical evaluation of ventilator-free days as an efficacy measure in clinical trials of treatments for acute respiratory distress syndrome. Crit Care Med. 2002;30(8):1772-7.
124. Granholm A, Anthon CT, Kjær M-BN, Maagaard M, Kaas-Hansen BS, Sivapalan P, et al. Patient-Important Outcomes Other Than Mortality in Contemporary ICU Trials: A Scoping Review. Critical Care Medicine. 2022;50(10):e759-e71.
125. European Comittee of Antimicrobial Susceptibility Testing (EUCAST), European Society of Clinical Microbiology and Infectious Diseases. Clinical breakpoints v. 13.1 - new file for printing. 29 June 2023. Accessed on 30 August 2023.
126. Leekha S, Terrell CL, Edson RS. General principles of antimicrobial therapy. Mayo Clin Proc. 2011;86(2):156-67.
127. Myles PS, Williamson E, Oakley J, Forbes A. Ethical and scientific considerations for patient enrollment into concurrent clinical trials. Trials. 2014;15:470.
128. Laine C, Taichman DB, Mulrow C. Trustworthy clinical guidelines. Ann Intern Med. 2011;154(11):774-5.
129. Qaseem A, Forland F, Macbeth F, Ollenschlager G, Phillips S, van der Wees P, et al. Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med. 2012;156(7):525-31.
130. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-6.
131. Burgess E, Singhal N, Amin H, McMillan DD, Devrome H. Consent for clinical research in the neonatal intensive care unit: a retrospective survey and a prospective study. Arch Dis Child Fetal Neonatal Ed. 2003;88(4):F280-5; discussion F5-6.
132. Faden RR, Beauchamp TL, Kass NE. Informed consent, comparative effectiveness, and learning health care. N Engl J Med. 2014;370(8):766-8.
133. Randolph AG. The unique challenges of enrolling patients into multiple clinical trials. Crit Care Med. 2009;37(1 Suppl):S107-11.
134. Cook D, McDonald E, Smith O, Zytaruk N, Heels-Ansdell D, Watpool I, et al. Co-enrollment of critically ill patients into multiple studies: patterns, predictors and consequences. Crit Care. 2013;17(1):R1.
135. Cook DJ, Ferguson ND, Hand L, Austin P, Zhou Q, Adhikari NK, et al. Coenrollment in a randomized trial of high-frequency oscillation: prevalence, patterns, predictors, and outcomes*. Crit Care Med. 2015;43(2):328-38.
136. Official Journal of the European Union. Regulation (EU) no 536/2014 of the European Parliament and of the council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC. 27 May 2014. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A32014R0536#d1e2403-1-1. Accessed on 25 February 2024.
137. Rothman KJ, Lash TL, Greenland S. Chapter 14: Introduction to categorical statistic. Modern Epidemiology (3rd ed). Lippincott Williams & Wilkins, 2012:237-257.
138. Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Medicine. 2018;16(1):29.

139. Gramacy RB. Chpater 5: Gaussian Process Regression. Gaussian Process Modeling, Design and Optimization for the Applied Sciences. Chapman Hall/CRC, Boca Raton, Florida, USA. 2020. Available from: https://bookdown.org/rbg/surrogates/chap5.html. Accessed on 4 April 2023.
140. Viele K, Saville BR, McGlothlin A, Broglio K. Comparison of response adaptive randomization features in multiarm clinical trials with control. Pharm Stat. 2020;19(5):602-12.

19 Appendices

19.1 Appendix 1: Charter for the independent data monitoring and safety committee

Introduction

The independent data monitoring and safety committee (IDMSC) will prepare a monitoring and meeting plan themselves. However, this charter defines the minimum of obligations and primary responsibilities of the IDMSC; its relationship with other trial components; its membership; and the purpose and timing of its meetings as perceived by the EMPRESS management committee. The charter also outlines the procedures for ensuring confidentiality and proper communication, the statistical monitoring guidelines and reports presented to the IDMSC, and an outline of the content of the open and closed reports which will be provided to the IDMSC.

Primary responsibilities of the IDMSC

The IDMSC are responsible for safeguarding the interests of trial participants, assessing the safety and efficacy of the interventions during the trial, and for monitoring the overall conduct of the trial. The IDMSC will provide non-binding recommendations about stopping or continuing the trial to the management committee of the EMPRESS trial. The IDMSC may also – if applicable - formulate recommendations related to the selection/recruitment/retention of participants, their management, adherence to protocol-specified regimens and retention of participants, and the procedures for data management and quality control.

The IDMSC will be advisory to the EMPRESS management committee. The management committee will be responsible for promptly reviewing the IDMSC recommendations, to decide whether to continue or terminate the trial, and to determine whether amendments to the protocol or changes in trial conduct are required.

Adaptive analyses and statistical methodology

Prior to commencement of randomisation, the IDMSC will receive detailed trial information, including the protocol, methodology, statistical approach, and adaptation rules in a physical, phone, or video meeting with at least one clinical member and one methodological or statistical member of the management committee. In addition, the IDMSC will receive a simulated dataset of similar structure as the datasets used during the adaptive analyses (including all relevant variables) and complete analysis code (in R) similar to

the code that will be used during all adaptive analyses (with the possibility that minor amendments will be required as described above, e.g., merging of small sites) prepared by the methodological/statistical centres of EMPRESS. The IDMSC statistician will be responsible for running and verifying this analysis code once; any questions or amendments will be resolved by contact with the EMPRESS methodological/statistical team, until the IDMSC statistician can vouch for the analysis code. This is done only once, as the IDMSC will not have the responsibility to run all adaptive (interim) analysis during the trial in duplicate due to their large number and the standardised procedure used for these analyses. Instead, the full IDMSC will be provided with the results in a format similar with mock **Table S3** in appendix 7, section 19.7 and information about whether a stopping rule has been reached as soon as possible after an adaptive analysis has been conducted. In addition, the IDMSC statistician will receive the dataset and code used to conduct the analysis, to make it possible for the IDMSC statistician to replicate the analyses if desired. As the adaptation rules are binding, the IDMSC will not have to decide upon whether the trial will continue based on the regular adaptive analyses. However, the IDMSC will have 48 hours after receiving the results of an adaptive analysis to object to the immediate implementation of the results (i.e., stopping the trial or changing the allocation probabilities) in case of any concerns regarding the analysis results. In that case, the IDMSC and sponsor will, as soon as possible, decide on a plan for the replication/verification of the analysis by the IDMSC (or other required steps) that will be implemented within the shortest possible time-frame in order to avoid unnecessary delays in adaptations.

Safety and yearly formal meetings

The IDMSC will receive additional unblinded safety data once yearly (i.e., descriptive data on SARs, including SUSARs, stratified by group), and will be required to meet at least once yearly, physically or by phone, after receiving safety data. In addition, the IDMSC plan additional meetings at their own discretion, and the IDMSC plans whether to contact each other by telephone, e-mail, or other means between meetings. At the yearly meetings, the IDMSC will be responsible of making a recommendation to the trial management committee about whether to continue, alter, pause, or stop the trial based on their review of safety data. It is the responsibility of the IDMSC to communicate their recommendation to the EMPRESS management committee without delay. Following a recommendation by the IDMSC, the EMPRESS management committee will, as fast as possible and within 48 hours, inform all trial sites and investigators about the recommendation of the IDMSC and the management committee's decision following this.
The IDMSC can, at any time during the trial, request additional information about outcome data, and may also request or conduct additional analyses, e.g., including external data from other trials that become available during trial conduct.

Members of the IDMSC

The IDMSC is an independent, multidisciplinary group consisting of a clinician, a trialist and a biostatistician that, collectively, has experience in the conduct, monitoring, and analysis of randomised clinical trials. The IDMSC will be thoroughly informed about the EMPRESS trial, including the adaptive methodology, prior to commencement of randomisation, as described above.

IDMSC trialist (chair)

Kathy Rowan, The Intensive Care National Audit and Resource Centre (ICNARC)
London, United Kingdom

IDMSC clinician

Lennie Derde, Department of Intensive Care Medicine, UMC Utrecht
Utrecht, The Netherlands

IDMSC statistician

Erin Evelyn Gabriel, Section of Biostatistics, Department of Public Health University of Copenhagen, Copenhagen, Denmark

Conflicts of interest

The members of the IDMSC will fill-in and sign a conflicts of interest form. IDMSC membership is restricted to individuals free of conflict of interest related to the current trial. The source of these conflicts may be financial, scientific, or regulatory in nature. Thus, neither trial investigators nor individuals employed by the sponsor or individuals with regulatory responsibilities for the trial products, will be members of the IDMSC. Furthermore, the IDMSC members will not own stocks in the companies having products being evaluated in the EMPRESS trial.

The IDMSC members will disclose to fellow members any consulting agreements or financial interests they have with the sponsor of the trial, or with other sponsors having products that are being evaluated or having products that are competitive with those being evaluated in the trial. The IDMSC will be responsible for deciding whether these consulting agreements or financial interests materially impact their objectivity. The IDMSC members will be responsible for advising fellow members of any changes in these consulting agreements and financial interests that occur during the trial. Any IDMSC members who develop significant conflicts of interest during the trial should resign from the IDMSC.

IDMSC membership is to be for the duration of the clinical trial. If any members leave the IDMSC during the trial, the management committee will appoint the replacement(s).

Final analysis meeting

The three members of the IDMSC will meet when 30-day follow-up data for all participants have been obtained, verified, and analysed following the stopping of the trial, regardless of whether this was due to a stopping rule being triggered or the maximum sample size being reached.

Proper communication

The IDMSC will have access to evolving information from the clinical trial regarding comparative results of efficacy and safety data aggregated by treatment group and may not share these data with any external parts or with any members of the trial management committee, investigators, clinicians, research staff, participants, or relatives without explicit written permission by the management committee. Only the members of the management committee and trial organisation involved in data collection, verification, and analysis will have access to the accumulating trial data, and these persons will not be permitted to include patients in the trial. The adaptive (interim) analysis results data will only be shared with the IDMSC and otherwise only seen by the statistical team conducting the analyses until the trial has ended; other members of the management committee and trial organisation will only be informed when an adaptive analysis has been conducted whether or not the trial continues or is stopped (and if stopped, the reason, without providing any actual numbers until follow-up for the primary outcome has ended).

Procedures will be implemented to ensure that proper communication is achieved between the IDMSC and the management committee. To provide a forum for exchange of information among various parties who share responsibility for the successful conduct of the trial, a format for open sessions and closed sessions will be implemented. The intent of this format is to enable the IDMSC to preserve confidentiality of the comparative efficacy results while at the same time providing opportunities for interaction between the IDMSC and others who have valuable insights into trial-related issues.

Closed sessions and reports

Sessions involving only IDMSC membership who generates the closed reports (called closed sessions) will be held yearly to allow discussion of confidential data from the clinical trial. IDMSC members will receive information corresponding to that presented in **Table S3** (appendix 7, section 19.7), summary data on SARs

(including SUSARs) in each group, and data on protocol adherence and recruitment. The IDMSC will receive these data approximately three days prior to the date of the meeting.

Closed reports will include analysis of the primary outcome (analysed by the EMPRESS statistical team, with analyses replicated by the IDMSC if they consider this necessary) and proportions of participants with SARs, including SUSARs, in each group. These closed reports will be prepared by the independent IDMSC biostatistician, with assistance from the trial data manager and trial statistical team as required. The closed reports should provide information that is accurate and be made at latest within one month from the date of the IDMSC meeting.

Open reports

Following each IDMSC meeting, the IDMSC will prepare an open report that contains information about whether a stopping rule has been triggered (and in that case, which stopping rule, without further details), data on the primary outcome and SARs (including SUSARs) aggregated for all participants (i.e., not separately in each group), data on recruitment, and data on protocol adherence (separately in each group), and any recommendations from the IDMSCs, with rationales as required. These will be shared with the trial sponsor and management committee.

Minutes of the IDMSC Meetings

The IDMSC will prepare minutes of their meetings. The closed minutes will describe the proceedings from all sessions of the IDMSC meeting, including the listing of recommendations by the IDMSC. These minutes may not be made available to anyone outside the IDMSC until the trial is stopped.

Recommendations to the management committee

All adaptive stopping rules are binding. The IDMSC will recommend pausing or stopping the trial if group-differences in SARs, including SUSARS, are concerning (with no pre-defined rules). If the recommendation is to stop the trial, the IDMSC will discuss and recommend on whether the final decision to stop the trial will be made after the analysis of all participants included at the time (including participants randomised after the analysis leading to the decision) or whether a moratorium shall take place (setting the trial at hold) in the further inclusion of participants during these extra analyses. If further analyses of the participants included after the analysis leading to the recommendation is recommended, the rules for finally recommending stopping should be specified with the recommendation. Furthermore, the IDMSC can recommend pausing or stopping the trial if continued conduct of the trial clearly compromises participant

safety. Similarly, if convincingly strong external evidence arises, the IDMSC may discuss this and make recommendations for trial pausing, stopping, or alterations, to safeguard the safety of participants. All recommendations from the IDMSC will be non-binding and advisory to the sponsor and management committee only.

Further IDMSC responsibilities

The management committee is jointly responsible with the IDMSC for safeguarding the interests of participants and for the conduct of the trial. Recommendations to amend the protocol or change the conduct of the trial made by the IDMSC will be considered and accepted or rejected by the management committee. The management committee will be responsible for deciding whether to continue, alter, pause, or stop the trial based on the IDMSC recommendations.

The IDMSC will be notified of all changes to the trial protocol (including analyses plans, if relevant) or conduct. The IDMSC concurrence will be sought on all substantive recommendations or changes to the protocol or trial conduct prior to their implementation.

After completion of the yearly IDMSC meetings, the recommendations from the IDMSC and the conclusion reached by the management committee will be submitted to the competent authorities.

After completion of the full analysis of outcomes at day 30, the IDMSC will make a recommendation to the management committee to submit a primary report on 30-day outcomes or await the 90-day outcomes.

19.2 Appendix 2: Reporting checklists

SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents* (www.spirit-statement.org)

Section/item	Item No	Description	Page No
Administrative information			
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym	1
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry	1
	2b	All items from the World Health Organization Trial Registration Data Set	1
Protocol version	3	Date and version identifier	1
Funding	4	Sources and types of financial, material, and other support	65
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors	1
	5b	Name and contact information for the trial sponsor	1, 7
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities	65
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)	1, 7-9
Introduction			
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	13-18
	6b	Explanation for choice of comparators	13-18
Objectives	7	Specific objectives or hypotheses	19
Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory)	20-23
Methods: Participants, interventions, and outcomes			
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained	7, 27

Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	24-25
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered	27-30
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease)	25, 27-31
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests)	31
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial	30
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	34-35
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure)	23
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations	20, 53-55
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size	21-22
Methods: Assignment of interventions (for controlled trials)			
Allocation:			
Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions	20-21, 52-53
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned	20-21
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions	20-21
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how	22-23
	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial	Not relevant

Methods: Data collection, management, and analysis

Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol	32, 34, 40, 64
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	23, 34, 99-100
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol	67
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	48-63
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)	59-63
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)	49-50, 61-62

Methods: Monitoring

Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed	9, 81-86
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial	50-52
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	36-39
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor	64

Ethics and dissemination

Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval	1, 18
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators)	66

Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)	40, 113-114
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable	67-68
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial	67-68
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site	68
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators	67
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation	65
Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions	66
	31b	Authorship eligibility guidelines and any intended use of professional writers	66
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code	67-68
Appendices			
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates	Separate documents
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable	None besides those collected as part of regular clinical practice

*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons “Attribution-NonCommercial-NoDerivs 3.0 Unported” license.

Consolidated Standards of Reporting Trials (CONSORT) - Adaptive designs CONSORT Extension (ACE) checklist

Section/Topic	Item no	Checklist item	Page no
Title and abstract	1a	Identification as a randomised trial in the title	20
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see ACE checklist for abstracts, Table 3)	4-6
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	13-18
	2b	Specific objectives or hypotheses	19
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	20-23
	3b« ǂ	Type of adaptive design used, with details of the pre-planned trial adaptations and the statistical information informing the adaptations	20-21, 48-53
	3c«3b ‡	Important changes to the design or methods after trial commencement (such as eligibility criteria) outside the scope of the pre-planned adaptive design features, with reasons	Not relevant at this stage
Participants	4a	Eligibility criteria for participants	24-25
	4b	Settings and locations where the data were collected	27
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	27-30
Outcomes	6a ‡	Completely define pre-specified primary and secondary outcome measures, including how and when they were assessed. Any other outcome measures used to inform pre-planned adaptations should be described with the rationale	34, 98-100
	6b ‡	Any unplanned changes to trial outcomes after the trial commenced, with reasons	Not relevant at this stage
Sample size and operating characteristics	7a ‡	How sample size and operating characteristics were determined	20, 53-55
	7b ‡‡	Pre-planned interim decision-making criteria to guide the trial adaptation process; whether decision-making criteria were binding or non-binding; pre-planned and actual timing and frequency of interim data looks to inform trial adaptations	48-53
Randomisation			
Sequence generation	8a	Method used to generate the random allocation sequence	20-21, 52-53
	8b ‡	Type of randomisation; details of any restriction (such as blocking and block size); any changes to the allocation rule after trial adaptation decisions; any pre-planned allocation rule or algorithm to update randomisation with timing and frequency of updates	20-21, 52-53
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	20-21
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	20-21, 52-53
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	Not done – 22-23
	11b	If relevant, description of the similarity of interventions	Not relevant
	11c ǂ	Measures to safeguard the confidentiality of interim information and minimise potential operational bias during the trial	22-23
Statistical methods	12a ‡	Statistical methods used to compare groups for primary and secondary outcomes, and any other outcomes used to make pre-planned adaptations	48-63
	12b« ǂ	For the implemented adaptive design features, statistical methods used to estimate	48-63

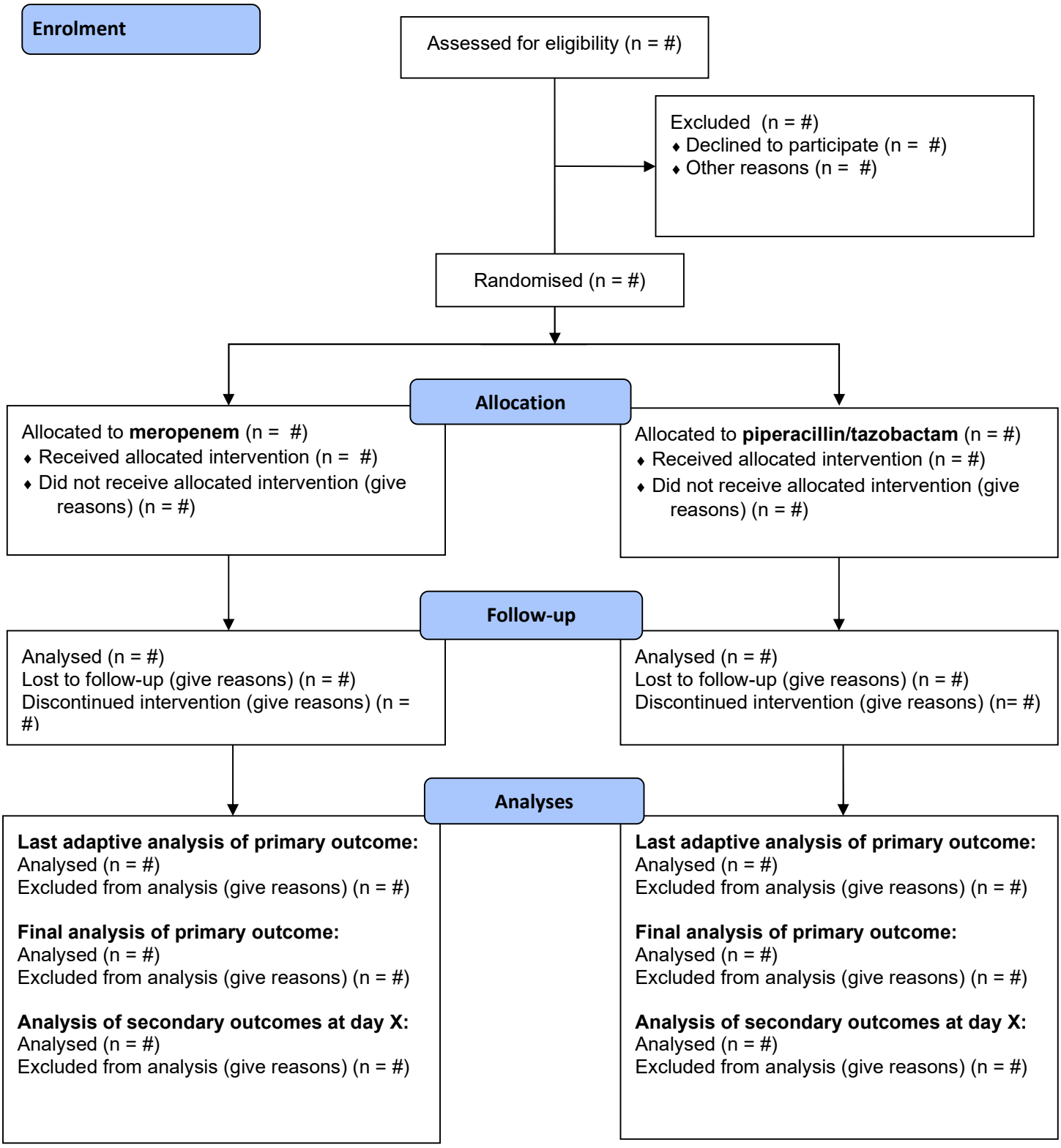
		treatment effects for key endpoints and to make inferences	
	12c«2b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	59-63
Results			
Participant flow (a diagram is strongly recommended)	13a ‡	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome and any other outcomes used to inform pre-planned adaptations, if applicable	94-95, 117-118
	13b	For each group, losses and exclusions after randomisation, together with reasons	94-95
Recruitment and adaptations	14a ‡	Dates defining the periods of recruitment and follow-up, for each group	Not relevant at this stage
	14b †	Why the trial ended or was stopped	50-52
	14c ǂ	Specify what trial adaptation decisions were made in light of the pre-planned decision-making criteria and observed accrued data	Not relevant at this stage
Baseline data	15a«15 †	A table showing baseline demographic and clinical characteristics for each group	115-116
	15b ǂ	Summary of data to enable the assessment of similarity in the trial population between interim stages	56
Numbers analysed	16 †	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	94-95, 117-118
Outcomes and estimation	17a †	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	117
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	117
	17c ǂ	Report interim results used to inform interim decision-making	118
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	59-63
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms) [1]	34, 36-39, 117
Discussion			
Limitations	20 †	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	Not relevant at this stage
Generalisability	21 †	Generalisability (external validity, applicability) of the trial findings	Not relevant at this stage
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	Not relevant at this stage
Other information			
Registration	23	Registration number and name of trial registry	1
Protocol	24a«24	Where the full trial protocol can be accessed	This document
SAP and other relevant trial documents	24b ǂ	Where the full statistical analysis plan and other relevant trial documents can be accessed	This document
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	65

SAP, statistical analysis plan; ACE, Adaptive designs
CONSORT Extension; "X« Y" means original
CONSORT 2010 item Y has been renumbered to X;
"X«" means item reordering resulted in new item X replacing the number of the original CONSORT 2010 item X
ǂ New items that should only be applied in reference to the ACE;
‡ Modified items that require reference to both CONSORT 2010 and ACE;
‡‡ Replacement (modified) item that only requires reference to the ACE;

† Item wording remains unchanged in reference to CONSORT 2010 but we expanded the ACE explanatory text to clarify additional considerations for certain adaptive designs. These unchanged items require reference to CONSORT 2010 except item 14b.

19.3 Appendix 3: Trial flowchart (adapted from the CONSORT 2010 flowchart)

Figure S1: Trial flowchart



Please refer to the CONSORT Statement for more information (http://www.consort-statement.org/) (122). The flowchart below has been adapted from the CONSORT flowchart and will be modified to reflect the flow of participants in the trial, including the number of participants at each stage.

19.4 Appendix 4: Definitions of variables collected in the EMPRESS trial

Definition of inclusion criteria

- Age ≥ 18 years: age of the participant in whole years at enrolment should be 18 years or above. The age of the participant will be calculated using the date of randomisation and the date of birth.
- Sepsis: Sepsis will be defined according to the Sepsis-3 criteria (1), i.e., suspected or documented infection and an acute increase of ≥ 2 points in the SOFA score (a proxy for organ dysfunction; if no known organ dysfunction or previous SOFA score available, it will assumed to be 0) (53).
- Septic shock: septic shock will be defined according to the Sepsis-3 criteria (1), i.e., sepsis and vasopressor requirement to maintain a mean arterial blood pressure of 65 mmHg or above and serum lactate levels greater than 2 mmol/L (53).
- Critical illness defined as use of at least one of the following:
    a. Invasive mechanical ventilation: use of mechanical ventilation via a cuffed endotracheal tube or tracheostomy at the time of randomisation.
    b. Non-invasive ventilation: non-invasive ventilation includes positive pressure ventilation via a tight mask or helmet
    c. CPAP for hypoxia: continuous use of CPAP via mask, helmet, or tracheostomy for hypoxia, i.e., not including intermittent use of CPAP.
    d. Oxygen supplementation with an oxygen flow of ≥ 10 litres/minute independent of delivery system: oxygen supplementation with an oxygen flow ≥10 L/min irrespectively of system used (mask or nasal cannula) or the addition of atmospheric air.
    e. Continuous infusion of any vasopressor or inotrope: infusion of any vasopressor/inotrope agent (i.e., norepinephrine, epinephrine, phenylephrine, vasopressin analogues, angiotensin, dopamine, dobutamine, milrinone or levosimendan) at the time of randomisation (excluding strictly procedure-related infusions).
- Clinical indication for empirical treatment with either meropenem or piperacillin/tazobactam: indication for use of empirical treatment with meropenem or piperacillin/tazobactam in the opinion of the treating clinician. Empirical antibiotic treatment is defined as use of antibiotics due to clinical suspicion of infection before culture and susceptibility results are available.

Definition of exclusion criteria

- Preceding intravenous treatment with meropenem or piperacillin/tazobactam for > 24 hours: use of at least one dose of meropenem or piperacillin/tazobactam on two consecutive days up to the day of randomisation.
- Fertile woman (< 60 years of age) with known pregnancy or positive urine hCG or plasma-hCG: Female participants under the age of 60 years with confirmed pregnancy by positive urine hCG or plasma-hCG.
- Known hypersensitivity or allergy to beta-lactam antibiotics: history of any hypersensitivity reaction (e.g., urticaria, eczema, angioedema, bronchospasm, and anaphylaxis) to any beta-lactam antibiotic (i.e., penicillins, cephalosporins, monobactams, carbapenems).
- Suspected or document central nervous system infection: clinical suspicion of or documented central nervous infection (i.e., meningitis, encephalitis, brain abscesses, subdural empyema or endophthalmitis).
- Known colonization or infection with bacteria with acquired resistance against meropenem or piperacillin/tazobactam within the previous 3 months (e.g., ESBL-, AmpC- or carbapenemase-producing bacteria): previous culture(s) and susceptibility tests showing pathogens with resistance against meropenem or piperacillin/tazobactam within the previous 3 months up to the date of randomisation. Vancomycin-resistant *Enterococci* (VRE) or methicillin-resistant *Staphylococcus aureus* (MRSA) will not lead to exclusion.
- Current or planned use of valproate within 30 days from randomisation: patients currently using or planned to use valproate in any dose, duration, and formulation within 30 days of randomisation.
- Patient included in another interventional trial where co-enrolment with EMPRESS is not permitted: patient enrolled in another interventional trial, where co-enrolment with EMPRESS is not permitted (i.e., if the two protocols collide or if the other trial does not allow co-enrolment with any other trial).
- Previously randomised into the EMPRESS trial: patients who has previously undergone randomisation in the EMPRESS trial at any time.
- Informed consent following inclusion expected to be unobtainable: patients where the clinician or investigator expects to be unable to obtain the necessary consent following inclusion according to the national regulations.
- Patient under coercive measures: patients with ongoing involuntary hospital admission or under the jurisdiction of correctional authorities.

Definition of clinical outcomes

- All-cause mortality at day 30, 90, and 180: death from any cause within 30, 90 or 180 days of randomisation.
- Number of participants with one or more of the following SARs within 30 of randomisation:
    - Anaphylactic shock to IV piperacillin/tazobactam or meropenem: anaphylactic reactions defined as urticarial skin reaction AND at least one of the following observed after randomisation in relation to skin reaction:
        - Worsened circulation (>20% decrease in blood pressure or new vasopressor infusion or >20% increase in vasopressor dose).
        - Increased airway resistance (>20% increase in the peak pressure on the ventilation).
        - Clinical stridor or bronchospasm.
        - Subsequent treatment with bronchodilators.
    - Invasive fungal infection: any of the following:
        - Suspected invasive fungal infection: presence of plasma markers in blood (e.g., candida mannan antigen and galactomannan antigen).
        - Confirmed invasive fungal infection: positive culture from blood, peritoneal fluid, or tissue.
    - Pseudomembranous colitis: symptoms and signs of infection in large intestine and a at least one culture or polymerase chain reaction (PCR) test demonstrating *Clostridioides difficile* in faeces.
    - Toxic epidermal necrolysis: confirmed or suspected (i.e., symptoms and signs of toxic epidermal necrolysis, e.g., extensive exfoliation of the epidermis and mucous membranes affecting more than 30% of the epidermis) toxic epidermal necrolysis according to treating clinicians (i.e., noted in electronic patient records).
- Number of participants with new isolation precautions due to one or more resistant bacteria within 30 days from randomisation: the number of participants who undergo new isolation precautions in hospital (i.e., new isolation precautions not already implemented at randomisation) due to colonisation or infection with one or more resistant bacteria within 30 days from randomisation. The resistant bacteria will be recorded according to the following groups:
    - Vancomycin-resistant *Enterococci* (VRE)
    - *Clostridioides difficile*

- Carbapenemase-producing bacteria, e.g., carbapenemase-producing *Enterobacterales, Acinetobacter,* or *Pseudomonas*
- ESBL-producing *Enterobacterales*
- AmpC β-lactamase-producing *Enterobacterales*
- *Citrobacter freundii*
- MRSA
- *Stenotrophomonas maltophilia*
- Other

- Days alive without life support (i.e., invasive mechanical ventilation, circulatory support, or renal replacement therapy [including days in between intermittent renal replacement therapy]) from randomisation to day 30 and day 90: the total number of days alive without the use of any of the following:
  - Invasive mechanical ventilation: defined under *Definition of the inclusion criteria.*
  - Circulatory support (for at least 1 hour on each day): defined under *Definition of inclusion criteria.*
  - Any form of renal replacement therapy (on each day): use of any form of in-hospital renal replacement therapy (e.g., dialysis, hemofiltration or hemodiafiltration) at any rate on each day, including days between intermittent renal replacement therapy for up to 3 days in between each renal replacement therapy. We will not register the use of dialysis at home.

  Participants who die during the follow-up period will be assigned zero days alive without life support as death is considered the worst possible outcome regardless of the duration of life support used. Moreover, this is in accordance with previous outcome definitions and is the most common way of handling death when using this type of outcome (123, 124).
- Days alive and out of hospital at day 30 and day 90: the days alive and not admitted to hospital within 30 or 90 days from randomisation. The number of days alive and out of hospital will be calculated using the discharge date from the index hospitalisation, the number of days readmitted to hospital (if any) and date of death, if relevant within the follow-up period. As for days alive without life support, non-survivors will be assigned zero days. Days in rehabilitation facilities, nursing homes, hospices, and other non-hospital institutions will not be considered as days in hospital.
- HRQoL at day 180 using EQ-5D-5L index values (57): HRQoL at 180 days (and up to 2 weeks from the time of follow up) will be assessed using the EQ-5D-5L questionnaire obtained by mail or phone as chosen by the participant. Non-survivors will be assigned the value 0, which corresponds to a

health state as bad as being dead (57). If the participant is incapable of answering the questionnaire (e.g., due to cognitive impairment or coma), we will ask the relative(s) to assess HRQoL for the trial participant using the proxy questionnaire. EQ-5D-5L will be converted to an index value in quantifying the self-rated health.

- HRQoL at day 180 using EQ VAS (57): HRQoL at 180 days (and up to 2 weeks from the time of follow up) will be assessed using the EQ VAS score (0-100) obtained by mail or phone as chosen by the participant. Non-survivors will be given the worst possible score (i.e., 0). If the participant is incapable of answering the questionnaire (e.g., due to cognitive impairment or coma), we will ask the relative(s) to assess HRQoL for the trial participant using the proxy questionnaire.

Feasibility outcomes

- Time to completion of feasibility phase: the time from randomisation of the first participant to randomisation of participant number 200 in months.
- Recruitment proportion: the number of randomised participants divided by the number of screened patients.
- Proportion of participants without consent to the use of data: the number of participants for whom the participant themselves or the relatives do not consent to the use of data divided by the number of randomised participants.
- Protocol adherence: the number of participants without one or more protocol violations (as defined in section 8.9) divided by the total number of randomised participants.
- Retention proportion: the number of participants with data on the primary outcome within maximum 15 days of follow up divided by the total number of randomised participants.

Screening variables

- Inclusion and exclusion criteria: defined under *Definition of inclusion criteria and Definition of exclusion criteria*
- Age at enrolment (date of birth)
- Presence of haematological or metastatic cancer (yes/no):
    - Haematological malignancy includes any of the following:
        - Leukaemia: Acute lymphoblastic leukaemia, acute myelogenous leukaemia, chronic myelogenous leukaemia, chronic lymphocytic leukaemia.

        - Lymphoma: Hodgkin's disease, and Non-Hodgkin lymphoma (e.g., small lymphocytic lymphoma, diffuse large B-cell lymphoma, follicular lymphoma, and mantle cell lymphoma).
        - Hairy cell leukaemia, marginal zone lymphoma, Burkitt's lymphoma, posttransplant lymphoproliferative disorder, T-cell prolymphocytic leukaemia, B-cell prolymphocytic leukaemia, Waldenström's macroglobulinemia and other NK- or T-cell lymphomas
        - Multiple myeloma/plasma cell myeloma
    - Metastatic cancer: proven metastasis by surgery, computed tomography (CT) scan, or any other method
- Acute surgical admission (yes/no): participants who have undergone any acute surgical procedure (i.e., surgery added to the operating room schedule 24 hours or less before surgery) during the current admission within 7 days prior to randomisation.
- Use of respiratory support at randomisation: as defined under *Definition of inclusion criteria*
- Latest FiO_2 prior to randomisation: only for participants on invasive mechanical ventilation, non-invasive ventilation, or continuous use of CPAP. The latest level of FiO_2 in percent measured before the time of randomisation.
- Maximum oxygen flow at randomisation: the maximum level of oxygen flow measured in l/min at the time of randomisation (+/- 1 hour), only for participants receiving supplementary oxygen on an open system. For participants using combinations of atmospheric air and pure oxygen, the total flow of oxygen will be calculated from the total flow of air and the FiO_2 using the following formula:
  Total flow of oxygen (l/min) = (total flow of air [l/min] * FiO_2/0.79) – (0.266 * total flow of air)
  If the FiO_2 is unknown, we will use the following formula to calculate the FiO_2 before using the formula above to calculate the total oxygen flow:
  FiO_2 = (flow of oxygen [l/min] + flow of atmospheric air [l/min] * 0.21) / (flow of oxygen [l/min] + flow of atmospheric air [l/min])
- Use of circulatory support (infusion of vasopressor/inotropes) at randomisation (yes/no): as defined under *Definition of inclusion criteria*
- Use of renal replacement therapy within the last 72 hours prior to randomisation (yes/no): use of any form of in-hospital renal replacement therapy (e.g., dialysis, hemofiltration, or hemodiafiltration) at any rate within the last 72 hours prior to randomisation. We will not register the use of dialysis at home.

- Trial site: all participating trial sites will be assigned a number identifying the site. If the same research unit randomises and follows participants across multiple departments, this will be considered a single site, but separately randomising departments in the same hospital will be considered as separate sites.

Baseline variables

- Sex: the sex of the participant assigned at birth (male or female)
- Coexisting conditions: any chronic co-morbidity present in the past medical history prior to randomisation and defined as follows:
    - History of ischaemic heart disease or heart failure (yes/no): history of ischaemic heart disease or heart failure: previous myocardial infarction, invasive intervention for coronary artery disease, stable or unstable angina, New York Heart Association functional class 3 or 4 or any measured left ventricular ejection fraction <40%.
    - Diabetes mellitus (yes/no): treatment at time of hospital admission with any anti-diabetic medications.
    - Chronic pulmonary disease (yes/no): treatment at time of hospital admission with any relevant drug indicating chronic pulmonary disease.
    - Known use of immunosuppressive therapy within the last 3 months: chronic use of systemic corticosteroids (excluding short-term use for e.g., exacerbations of pulmonary disease) or other systemic immunosuppressive drugs (e.g., tumour necrosis factor inhibitors, calcineurin inhibitors, mTOR inhibitors, anti-thymocyte globulins, interleukin (IL) 2 inhibitors, IL-6 inhibitors, mycophenolate, azathioprine, belimumab) or chemotherapy (e.g., alkylating agents, anti-metabolites, mitotic inhibitors, topoisomerase inhibitors, others) within the last 3 months before randomisation.
    - Previous organ transplantation: history of any solid organ transplantation except for cornea.
    - Chronic liver disease: portal hypertension; cirrhosis proven by biopsy, computed tomography (CT) scan or ultrasound; history of variceal bleeding; or hepatic encephalopathy in the past medical history.
    - Chronic use of renal replacement therapy: hronic use of any form of RRT (e.g., haemodialysis or peritoneal dialysis) at least one a week prior to index hospital admission.
- Clinical Frailty Scale, version 2.0 (level 1-9) (63, 64): investigator-assessed clinical frailty scale. The investigator will assign the participant one value from 1 (very fit) to 9 (terminally ill) representing

the measure of fitness and frailty (63, 64) based on information from the participants, the relatives, and the electronic patient records.

- Date of admission to hospital: the date of admission to the first hospital the participant was admitted to during the current hospital admission
- Type of department at which the participant was included: the type of department that the participants was admitted to at the time of randomisation.
    - Emergency department: accident/emergency/casualty/acute department
    - Hospital ward: medical or surgical ward
    - Intermediate care unit: area of the hospital with higher resources to monitor patients as defined by the site, but invasive mechanical cannot be given.
    - ICU: area of the hospital where invasive mechanical ventilation can be given.
- Primary site of infection (suspected or confirmed): the primary anatomical site of infection defined as one of the following:
    - Pulmonary: any infection located in the lungs, including pneumonia, lung abscess, and empyema
    - Gastrointestinal: any infection located in the gastrointestinal tract, including peritonitis, abscess, cholangitis, cholecystitis, diverticulitis, and diarrheal diseases
    - Urinary tract: any infection located in the urinary tract, including pyelonephritis and cystitis.
    - Skin or soft tissue: any infection located in the skin or soft tissues, including cellulitis, phlegmon, erysipelas, or fasciitis
    - Bloodstream: any infection located in the bloodstream, including endocarditis, central venous catheter-associated bloodstream infections and bacteraemia
    - Other infections, including unknown primary site of infection: any other infection not mentioned in the subcategories above or infection without any confirmed or suspected primary focus.
- Positive bacterial cultures from specific sites in the 48 hours prior to and 1 hour after randomisation (yes/no): any culture showing the presence of one or more bacteria in a sample taken from usually sites in the 48 hours before and up to 1 hour after randomisation (see '*Sample type*' below). For each positive culture, we will record the type of sample, bacteria, and the sensitivity to both trial drugs:
    - Type of sample:
        - Blood

        - Lower airway secretions or bronchoalveolar lavage
        - Urine
        - Tissue, bone, or pus
        - Peritoneal fluid
        - Pleural fluid
        - Cerebrospinal fluid
    - Type of bacteria (for each sample):
        - *Achromobacter* spp.
        - *Acinetobacter baumanii*
        - *Aerococcus* spp.
        - *Aeromonas* spp.
        - Anaerobes
        - *Burkholderia* spp.
        - *Campylobacter* spp.
        - *Citrobacter* spp.
        - *Enterobacter* spp.
        - *Enterococcus faecalis*
        - *Enterococcus faecium*
        - Other *Enterococcus*
        - *Escherichia coli*
        - *Haemophilus influenza*
        - *Klebsiella* spp.
        - *Legionella* spp.
        - *Listeria* spp.
        - *Moraxella catarrhalis*
        - *Neisseria* spp.
        - *Proteus* spp.
        - *Pseudomonas* spp.
        - *Salmonella* spp.
        - *Serratia* spp.
        - *Shigella* spp.
        - Methicillin-resistant *Staphylococcus aureus (*MRSA)
        - Methicillin-sensitive *Staphylococcus aureus (*MSSA)

        - *Stenotrophomonas* spp.
        - *Streptococcus pneumoniae*
        - Other *Streptococcus*
        - Other bacteria
    - Resistance to trial drugs (for each bacterium):
        - Known resistance to piperacillin/tazobactam (yes/no), i.e., resistant as per EUCAST (125)
        - Known resistance to meropenem (yes/no), i.e., resistant as per EUCAST (125)
- Limitations of care (yes/no): participant with limitation(s) in use of life support (i.e., invasive mechanical ventilation, circulatory support, renal replacement therapy) and/or cardio-pulmonary resuscitation at the time of randomisation.
- Treatment during current hospital admission prior to randomisation:
    - Anti-bacterial agent (yes/no): any antibiotic treatment (IV, intramuscular, oral or per gastrointestinal tube; not including transdermal antibiotics) commenced due to suspected bacterial infection before microbiological results are available
        - Beta-lactamase/beta-lactamase inhibitor, e.g., piperacillin/tazobactam
        - Carbapenem, e.g., meropenem
        - Cephalosporin, e.g., cefuroxime, ceftriaxone
        - Penicillin
        - Glycopeptide, e.g., vancomycin
        - Fluoroquinolone, e.g., ciprofloxacin
        - Macrolide, e.g., clarithromycin
        - Clindamycin
        - Nitroimidazole, e.g., metronidazole
        - Aminoglycoside, e.g., gentamycin
        - Oxazolidinone, e.g., linezolid
        - Sulfonamide, e.g., sulfametizole
        - Tetracycline, e.g., doxycycline, tigecycline
        - Polymyxin, e.g., colistin
        - Other: any other antibiotic agent not included in categories above
- Participant weight (kg): estimated or measured.
- Participant height (m): estimated or measured.

- PaO_2 and SaO_2: PaO_2 and SaO_2 will be assessed from the most recent arterial blood gas sample. Alternatively, if an arterial blood gas sample is not available, SpO_2 will be assessed from the most recent measure by pulse oximeter up to the time of randomisation.
- Lowest registered systolic blood pressure in the 24 hours preceding randomisation: lowest systolic blood pressure measured either invasively or non-invasively in mmHg. In case of cardiac arrest in the 24 hours preceding randomisation, 0 mmHg will be registered.
- Highest plasma lactate in the 24 hours prior to randomisation (mmol/L): the highest level of lactate from any blood sample (i.e., venous, or arterial) in the 24 hours preceding randomisation.
- Highest plasma creatinine in the 24 hours prior to randomisation (µmol/L): the highest level of creatinine from any blood sample (i.e., venous, or arterial) in the 24 hours preceding randomisation.

Daily during admission for the first 30 days after randomisation (dayforms)

- Death (yes/no): defined under *Definition of clinical outcomes.*
- Admitted to hospital (yes/no): defined under *Definition of clinical outcomes*.
- Use of invasive mechanical ventilation (yes/no): defined under *Definition of inclusion criteria.*
- Use of circulatory support (continuous infusion of vasopressor/inotropes for a minimum of 1 hour) on this day (yes/no): defined under *Definition of inclusion criteria.*
- Use of any form of renal replacement therapy on this day including days between intermittent renal replacement therapy (yes/no): defined under *Definition of clinical outcomes*
- SAR(s) on this day (yes/no for each): defined under *Definition of clinical outcomes*
    - Anaphylactic shock to IV piperacillin/tazobactam or meropenem: defined under *Definition of clinical outcomes*
    - Invasive fungal infection: defined under *Definition of clinical outcomes*
    - Pseudomembranous colitis: defined under *Definition of clinical outcomes*
    - Toxic epidermal necrolysis: defined under *Definition of clinical outcomes*
- New isolation precautions due to resistant bacteria on this day (yes/no): any isolation precautions, which were not present at the time of randomisation, and which were implemented to prevent transmission of bacterial infectious diseases.
    - If yes, type of resistant bacteria:
        - Vancomycin-resistant *Enterococci*
        - *Clostridiodies difficile*
        - Carbapenemase-producing bacteria

        - Extended spectrum β-lactamase (ESBL) producing *Enterobacterales*
        - AmpC β-lactamase-producing *Enterobacterales*
        - *Citrobacter freundii*
        - Methicillin-resistant *Staphylococcus aureus* (MRSA)
        - *Stenotrophomonas maltophilia*
        - Other

Daily registration of co-interventions and protocol adherence up to day 30 after randomisation (only during intervention period)

- Trial intervention (yes/no): daily registration of whether the participant received at least 50% of the planned trial medication on each day as per protocol.
    - If no, apply reason:
        - Change to definitive treatment
        - De-escalation to another empirical antibiotic with more narrow spectrum due to clinical improvement (not including the two trial interventions), i.e., in situations where no bacteria are found in microbiological cultures
        - Empirical or definitive antibiotic treatment no longer indicated (including switching to prophylactic antibiotic treatment)
        - Clinical deterioration with indication for different antibiotic than allocated according to treating clinician
        - Adverse reaction to the allocated agent requiring shift of antibiotic according to treating clinician
        - By error
        - Other reason(s)
- Use of other antibiotic agent than allocated on this day (yes/no): use of any other prophylactic, empirical or definitive antibiotic treatment (IV, oral or per gastrointestinal tube; not including transdermal antibiotics) commenced due to suspected or confirmed bacterial infection.
    - If yes, apply type:
        - Prophylactic treatment: antibiotic treatment used in patients at high risk of but without clinical signs or symptoms of infection (126).
        - Empirical treatment: antibiotic treatment used in patients with suspected infection before results of cultures and susceptibility tests are available (126).

        - Definitive treatment: antibiotic treatment used in patients with documented infection according to the results of cultures and susceptibility tests (126).
    - If yes, apply type of agent
        - Beta-lactamase/beta-lactamase inhibitor, e.g., piperacillin/tazobactam
        - Carbapenem, e.g., meropenem
        - Cephalosporin, e.g., cefuroxime, ceftriaxone
        - Penicillin
        - Glycopeptide, e.g., vancomycin
        - Fluoroquinolone, e.g., ciprofloxacin
        - Macrolide, e.g., clarithromycin
        - Clindamycin
        - Nitroimidazole, e.g., metronidazole
        - Aminoglycoside, e.g., gentamycin
        - Oxazolidinone, e.g., linezolid
        - Sulfonamide, e.g., sulfametizole
        - Tetracycline, e.g., doxycycline, tigecycline
        - Polymyxin, e.g., colistin
        - Other: any other antibiotic agent not included in categories above

Daily during days 31-90 after randomisation

- Death (yes/no): defined under *Definition of clinical outcomes.*
- Admitted to hospital (yes/no): defined under *Definition of clinical outcomes*.
- Use of invasive mechanical ventilation from day 31-90 (yes/no): defined under *Definition of inclusion criteria.*
- Use of circulatory support (continuous infusion of vasopressor/inotropes for a minimum of 1 hour) from day 31-90 (yes/no): defined under *Definition of inclusion criteria.*
- Use of renal replacement therapy from day 31-90, including days between intermittent renal replacement therapy (yes/no): defined under *Definition of clinical outcomes.*

Follow-up 180 days after randomisation

- Death (yes/no): defined under *Definition of clinical outcomes.*
- HRQoL assessed by EQ-5D-5L domains (level 1, 2, 3, 4 or 5 for each domain): defined under *Definition of clinical outcomes.*

- HRQoL assessed by EQ VAS (0-100): defined under *Definition of clinical outcomes.*
- Date of HRQoL assessment: the date of HRQoL survey of the participant or proxy. For non-responders, the reason(s) for not responding is collected using the following categorisation:
    - Do not wish to participate.
    - Death after day 180 but before HRQoL follow up was conducted.
    - Contact to participant or proxy could not be established.
    - Other.
- HRQoL assessor: registration if the HRQoL outcomes were collected by survey of the participant or proxy

Definitions of baseline covariates used in HTE analyses

These baseline covariates will also be included in the baseline data table.

- Disease severity according to SMS-ICU (109): 0-42 points, with higher scores indicating higher severity of illness and higher risk of death. Includes seven baseline covariates registered at the time of randomisation or in the 24 hours preceding randomisation: age (years), lowest systolic blood pressure (mmHg), haematological malignancy or metastatic cancer, acute surgical admission, use of vasopressors/inotropes, use of respiratory support, and use of renal replacement therapy (here, renal replacement therapy within the 72 hours preceding randomisation will be used). Variables are defined as elsewhere (109) and in this section.
- Primary site of infection: participants with pulmonary versus gastrointestinal versus urinary tract versus other primary sites of infections. Pulmonary, gastrointestinal, and urinary tract infections are defined under *Baseline variables,* whereas other infections encompass any other primary sites of infection.
- Immunosuppression: participants with versus without immunosuppression, which we define as participants with any of the following:
    - Presence of haematological cancer or metastatic cancer as defined under *Screening variables*.
    - Known use of immunosuppressive therapy within the last 3-months as defined under *Baseline variables*.
    - Previous organ transplantation as defined under *Baseline variables*.
- Creatinine: highest plasma creatinine value in µmol/L within the last 24 hours of randomisation.

19.5 Appendix 5: Co-enrolment

Based upon an updated critical appraisal of the literature, the EMPRESS management committee endorses and encourages co-enrolment in the EMPRESS trial. The following issues have been considered.

Ethical considerations

Preventing eligible patients from co-enrolment in trials, which they would potentially benefit from participating in, and whose material risks and benefits they understand, violates their autonomy - and thus contravenes a fundamental principle of research ethics (127).

Permitting co-enrolment is in accordance with existing recommendations for the conduct of trustworthy clinical practice guidelines taking into account benefits and harms, quality of evidence, values and preferences (of patients or their relatives) and cost considerations as outlined by the Institute of Medicine, the Guideline International Network, and according to The Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology (128-130).

Relatives have limited concerns about co-enrolment (131).

General considerations

Critically ill patients receive many different interventions in addition to the trial intervention due to acute and chronic illnesses. Consequently, the potential for interactions is a prerequisite in clinical trials in critically ill patients, and co-enrolment is thus not different from what occurs in single-enrolment trials (127).

In pragmatic trials (like the EMPRESS trial), other interventions will be given at random and are therefore difficult to control for. If interaction in fact is an issue, it may be better controlled for if patients are co-enrolled and randomised to more than one intervention.

Clinical research with a potential to inform and improve clinical practice is valuable and should be supported. More high-quality clinical research can be conducted in a timely fashion and more information can be generated to guide clinical practice, if co-enrolment is permitted (132).

Scientific and statistical considerations

Pragmatic clinical trials allowing inclusion of a broad range of trial participants and options for drug treatments and other therapies (co-enrolment) have higher external validity/generalisability than non-pragmatic trials with restrictions regarding trial participants and co-enrolment (133).

Non-pragmatic trials with restrictions regarding study participants and co-enrolment are exposed to drugs and other treatments in a less clinically relevant setting where interactions are largely uncontrolled and poorly evaluated. Co-enrolment in pragmatic trials facilitates evaluation of clinically relevant and patient-important interactions (127).

Co-enrolment into two or more trials does not invalidate the original randomisation of the individual trials unless the two protocols collide. Separate analysis of each individual trial, ignoring the issue of co-enrolment into the other trial, will retain the balance of participant characteristics expected by standard random assignment within each trial (127). While co-enrolment may affect power if a treatment in one trial affects outcomes (e.g., if two two-arm trials have mortality as the primary outcome and co-enrol, an intervention reducing mortality in one trial may lower power of the other due to overall fewer events), this issue is mitigated in trials with adaptive sample sizes, which in those cases may just include more patients than originally expected.

The National Institute of Health supports co-enrolment (133); so does the Canadian Critical Care Trials group (http://www.ccctg.ca/Home.aspx) and the Australian New Zealand Intensive Care Society's Clinical Trial Group (http://www.anzics.com.au/Pages/CTG/CTG-home.aspx).

Co-enrolment into two or more trials does not seem to affect the natural course of the disease of the other condition being studied (127). Co-enrolment does not appear to influence patient safety or trial results (134, 135). Empirically, co-enrolment has a small effect on study power (127).

In conclusion, we highly support and encourage co-enrolment because of overall benefit, including ethical, practical, and scientific benefit, and no evidence of harm.

Co-enrolment agreement form

We will encourage engagement in research projects other than the EMPRESS trial.

Please fill in the information of the trial to be evaluated as counterpart for co-enrolment with the EMPRESS trial and send it by e-mail to empress@cric.nu.

Once we have received the information below, we will contact the principal/coordinating investigator of the trial and facilitate exchange of protocols and other relevant documents between the management committees.

We have prepared the form for only one trial, but please feel free to copy as many forms as you need.

Official full/short title of the project:

Short description of the project:

Contact information of principal/coordinating investigator of the trial:

Name:

E-mail:

19.6 Appendix 6: Justification for the use of emergency situation clinical trial procedures

All patients eligible for enrolment in EMPRESS will be critically ill (due to sepsis or septic shock with severe hypoxia or hypotension), and the informed consent procedure will follow the applicable legislation for clinical trials conducted in emergency situations.

Consequently, enrolment will be without prior informed consent from the patients as critical illness due to sepsis constitutes an emergency situation, according to the criteria outlined in the EU regulation on clinical trials on medicinal products for human use (No 536/2014) (136):

- Critical illness due to sepsis is a sudden life-threatening condition, requiring urgent treatment with empirical broad-spectrum antibiotics as assessed in EMPRESS (62).
- Participation in the trial has the potential to directly benefit subjects by providing urgent treatment with one of the assessed interventions, in accordance with standard clinical care and international recommendations (62). As piperacillin/tazobactam is currently used more frequently than carbapenems (6, 7), the trial increases the probability that participants are treated with meropenem compared to usual clinical practice, and this is hypothesized to lead to improved outcomes. Finally, by using response-adaptive randomisation, the trial directly increases participants' chances of being allocated to the intervention with the highest probability of being superior (25).
- It is not possible within the therapeutic window to supply all prior information and obtain prior informed consent from the patient's proxy or legally designated representative, as immediate administration of empirical broad-spectrum antibiotics at the time of diagnosis is recommended (62) and delays in administration may negatively affect patient outcomes.
- If the investigator is aware of any objections to participate in the trial previously expressed by the patient, the patient will be excluded under the *"Informed consent following inclusion expected to be unobtainable"* exclusion criterion (section 7.2).
- The trial directly relates to the patient's medical condition because of which it is not possible to obtain prior informed consent from the patient or the patient's proxy/legally designated representative within the therapeutic window, and the trial is of a nature that it can only be conducted in emergency situations, as the clinical condition under study will always constitute a medical emergency.
- The trial poses minimal risk to and burden on the patient in comparison with standard treatment of the patient, as both interventions are already marketed, have well-known safety profiles, and are

currently regarded as part of standard of care for this condition, as described in detail in section 10.2.

A detailed description of the recruitment and informed consent procedure in each participating country will be submitted for approval by the relevant competent authorities and made available on the trial website prior to start of enrolment in each participating country. The procedure of enrolment without prior informed consent used in Denmark and the EU will be adapted for participating countries outside the EU if required by the relevant competent authorities.

19.7 Appendix 7: Mock baseline and outcome tables

Table S1: Baseline characteristics

Characteristic	Meropenem (N = #)	Piperacillin/tazobactam (N = #)
Country of enrolment		
Denmark	n (#.#%)	n (#.#%)
[Each additional participating country listed separately in the trial report(s)]	n (#.#%)	n (#.#%)
	n (#.#%)	n (#.#%)
	n (#.#%)	n (#.#%)
Age, median (IQR), years	## (## to ##)	## (## to ##)
Sex		
Female	n (#.#%)	n (#.#%)
Male	n (#.#%)	n (#.#%)
Weight, median (IQR), kg	## (## to ##)	## (## to ##)
Height, median (IQR), m	## (## to ##)	## (## to ##)
Coexisting conditions		
Ischemic heart disease or heart failure	n (#.#%)	n (#.#%)
Diabetes mellitus	n (#.#%)	n (#.#%)
Chronic pulmonary disease	n (#.#%)	n (#.#%)
Known immunosuppression	n (#.#%)	n (#.#%)
Haematological or metastatic cancer	n (#.#%)	n (#.#%)
Known use of immunosuppressive therapy within the last 3-months	n (#.#%)	n (#.#%)
Previous organ transplantation	n (#.#%)	n (#.#%)
Chronic use of renal replacement therapy	n (#.#%)	n (#.#%)
Clinical Frailty Scale, median (IQR) (63, 64)	# (# to #)	# (# to #)
Limitations of care	n (#.#%)	n (#.#%)
Time from hospitalization to enrolment, median (IQR), days	# (# to #)	# (# to #)
SMS-ICU, median (IQR) (109)	## (## to ##)	## (## to ##)
Place of enrolment		
Emergency department	n (#.#%)	n (#.#%)
Hospital ward	n (#.#%)	n (#.#%)
Intermediate care unit	n (#.#%)	n (#.#%)
Intensive care unit	n (#.#%)	n (#.#%)
Primary site of infection		
Pulmonary	n (#.#%)	n (#.#%)
Gastrointestinal	n (#.#%)	n (#.#%)
Urinary tract	n (#.#%)	n (#.#%)
Skin or soft tissue	n (#.#%)	n (#.#%)
Bloodstream	n (#.#%)	n (#.#%)
Other	n (#.#%)	n (#.#%)
Acute surgical admission	n (#.#%)	n (#.#%)
Oxygen supplementation		
Invasive mechanical ventilation	n (#.#%)	n (#.#%)
FiO_2, median (IQR), %	## (## to ##)	## (## to ##)
Non-invasive ventilation or continuous use of CPAP	n (#.#%)	n (#.#%)
FiO_2, median (IQR), %	## (## to ##)	## (## to ##)
Open system	n (#.#%)	n (#.#%)
O_2-flow, median (IQR), L/min	## (## to ##)	## (## to ##)
PaO_2, median (IQR), mmHg	## (## to ###)	## (## to ###)
SaO_2, median (IQR), %	## (## to ###)	## (## to ###)
Lowest systolic blood pressure, median (IQR), mmHg	### (### to ###)	### (### to ###)

Highest lactate, median (IQR), mmol/L	#.# (#.# to #.#)	#.# (#.# to #.#)
Highest creatinine, median (IQR), µmol/L	#.# (#.# to #.#)	#.# (#.# to #.#)
Cointerventions used at randomisation		
Systemic corticosteroids	n (#.#%)	n (#.#%)
Vasopressors or inotropes	n (#.#%)	n (#.#%)
Rrenal replacement therapy	n (#.#%)	n (#.#%)
Anti-bacterial agents		
Beta-lactamase/beta-lactamase inhibitor	n (#.#%)	n (#.#%)
Carbapenem	n (#.#%)	n (#.#%)
Cephalosporin	n (#.#%)	n (#.#%)
Penicillin	n (#.#%)	n (#.#%)
Glycopeptide	n (#.#%)	n (#.#%)
Fluoroquinolone	n (#.#%)	n (#.#%)
Macrolide	n (#.#%)	n (#.#%)
Clindamycin	n (#.#%)	n (#.#%)
Nitroimidazole	n (#.#%)	n (#.#%)
Aminoglycoside	n (#.#%)	n (#.#%)
Oxazolidinone	n (#.#%)	n (#.#%)
Sulphonamide	n (#.#%)	n (#.#%)
Tetracycline	n (#.#%)	n (#.#%)
Polymyxin	n (#.#%)	n (#.#%)
Other	n (#.#%)	n (#.#%)

Abbreviations (in alphabetical order): CPAP: continuous positive airway pressure, IQR: interquartile range, l/min: litres per minute, mmHg: millimetres of mercury, mmol/L: millimoles per litre, n: number, SMS-ICU: Simplified Mortality Score for the Intensive Care Unit (109), µmol/L: micromoles per litre.

Binary or categorical variables are presented as numbers with percentages; numerical variables are presented as medians with interquartile ranges. Definitions of all baseline variables are provided in appendix 4, section 19.4. The number and proportions of participants with missing data for all variables will be reported in the final table.

Table S2: Outcome data

	Descriptive outcome data		Treatment effect estimates		Probabilities of effects with meropenem	
Outcome	**Meropenem** (N = #)	**Piperacillin/ tazobactam** (N = #)	**Relative difference (RR/RoM, 95% CrI)**	**Absolute difference (RD/MD, 95% CrI)**	**Any benefit**	**Any harm**
All-cause mortality at day 30	n/N (#.#%)	n/N (#.#)	#.## (#.## to #.##)	#.#%-points (#.# to #.#)	#.##%	#.##%
One or more SARs at day 30	n/N (#.#%)	n/N (#.#%)	#.## (#.## to #.##)	#.#%-points (#.# to #.#)	#.##%	#.##%
New isolation precautions due to one or more resistant bacteria at day 30	n/N (#.#%)	n/N (#.#%)	#.## (#.## to #.##)	#.#%-points (#.# to #.#)	#.##%	#.##%
Days alive without life support at day 30	#.## (#.## to #.##)	#.## (#.## to #.##)	#.## (#.## to #.##)	#.## (#.## to #.##)	#.##%	#.##%
Days alive and out of hospital at day 30	#.## (#.## to #.##)	#.## (#.## to #.##)	#.## (#.## to #.##)	#.## (#.## to #.##)	#.##%	#.##%
All-cause mortality at day 90	n/N (#.#%)	n/N (#.#%)	#.## (#.## to #.##)	#.#%-points (#.# to #.#)	#.##%	#.##%
Days alive without life support at day 90	#.## (#.## to #.##)	#.## (#.## to #.##)	#.## (#.## to #.##)	#.## (#.## to #.##)	#.##%	#.##%
Days alive and out of hospital at day 90	#.## (#.## to #.##)	#.## (#.## to #.##)	#.## (#.## to #.##)	#.## (#.## to #.##)	#.##%	#.##%
All-cause mortality at day 180	n/N (#.#%)	n/N (#.#%)	#.## (#.## to #.##)	#.#%-points (#.# to #.#)	#.##%	#.##%
EQ-5D-5L index values at day 180	#.## (#.## to #.##)	#.## (#.## to #.##)	#.## (#.## to #.##)	#.## (#.## to #.##)	#.##%	#.##%
EQ VAS at day 180	#.## (#.## to #.##)	#.## (#.## to #.##)	#.## (#.## to #.##)	#.## (#.## to #.##)	#.##%	#.##%

Mock version of the primary table with data from the final, primary analyses of all clinical outcomes.

Abbreviations (in alphabetical order): CrI: credible interval; EQ VAS: EuroQol Visual Analogue Scale; MD: mean difference; n and N: n denotes number of participants with the outcome, N denotes to the total number of participants in a group; RD: (absolute) risk difference; RoM: ratio of means; RR: relative risk; SARs: serious adverse reactions.

Binary outcomes are presented as numbers with percentages; numerical outcomes are presented as medians with interquartile ranges.

Treatment effect estimates and probabilities of benefit/harm with meropenem are based on the posterior distributions of adjusted average treatment effects. For the primary outcome, 30-day mortality, probabilities of effect sizes smaller than the threshold for practical equivalence (an absolute risk difference of 2.5%-points) will also be presented along with the probabilities of treatment effects equal to or larger than 2.5%-points in both directions. For all outcomes, the complete posterior distributions of average treatment effects and the corresponding probabilities of *all* effect sizes will be visualised as previously done (90).

Table S3: Reporting of results for 30-day mortality from the adaptive analyses

	Analysed/ randomised		Descriptive outcome data		Treatment effect estimates		Probabilities of various effects with meropenem					Subsequent allocation ratios	
Adaptive analysis (number, last inclusion date, analysis date, implementation date)	**Mero-penem**	**Pipera-cillin/ tazobactam**	**Mero-penem**	**Pipera-cillin/ tazo-bactam**	**Absolute difference (RD, 95% CrI)**	**Relative difference (RR, 95% CrI)**	**Any benefit** (superiority with meropenem)	**Any harm** (inferiority with meropenem)	**Absolute difference > -2.5%-points and < 2.5%-points** (practical equivalence)	**Absolute difference ≤ -2.5%-points** (effect larger than practical equivalence threshold in beneficial direction)	**Absolute difference ≥ 2.5%-points** (effect larger than practical equivalence threshold in harmful direction)	**Mero-penem**	**Pipera-cillin/ tazobactam**
#1, YYYY-MM-DD, YYYY-MM-DD, YYYY-MM-DD	n/N (#.#%) [complete: n/N (#.#%)]	n/N (#.#%) [complete: n/N (#.#%)]	n/N (#.#%)	n/N (#.#%)	#.## (#.## to #.##)	#.# %-points (#.# to #.#)	#.##%	#.##%	#.##%	#.##%	#.##%	#.##%	#.##%
….													

Mock version of the table that will be used to present the results of all adaptive analyses of the primary outcome conducted during the trial; all adaptive analyses will be presented in a separate row in the table. If data completeness (of all variables included in the analyses) is not 100% in both groups, all results (descriptive and inferential) will be calculated after multiple imputation.

Abbreviations (in alphabetical order): CrI: credible interval; YYYY-MM-DD: year, month, date of the time of the follow-up dates and the dates where analyses were conducted and results implemented (e.g., allocation ratios changed); MD: mean difference; n and N: n denotes number of participants with the outcome, N denotes to the total number of participants in a group; RD: (absolute) risk difference; RoM: ratio of means; RR: relative risk. Numerical data are presented as numbers with percentages. Treatment effect estimates and probabilities of benefit/harm with meropenem are based on the posterior distributions of adjusted average treatment effects.

19.8 Appendix 8: Priors used in EMPRESS

This section outlines the prior probability distributions used for the analyses of all clinical outcomes in EMPRESS. All models used will have the following form and will be adjusted for the following baseline variables (section 12.7):

outcome ~ intercept + intervention + age + age squared + presence of haematological or metastatic cancer + acute surgical admission + use of invasive mechanical ventilation + use of circulatory support (continuous infusion of vasopressors/inotropes) + use of renal replacement therapy within the last 72 hours prior to randomisation + site of inclusion + time period

In addition, the linear models used for analysing continuous outcomes will also include an auxiliary nuisance parameter, *sigma*, corresponding to the estimated standard deviation (SD) of the residuals. All model terms will be on the natural scale without centring; appropriate error distributions and link functions will be used for the outcomes as described in section 12.7. Generally, the primary priors for the treatment effects will be neutral and slightly sceptical (to provide some regularisation against extreme fluctuations early in the trial (75)) and priors for covariates and nuisance parameters will be very vague; as such, the primary priors for everything but the treatment effect will generally be overwhelmed by the data accumulated at all planned adaptive analyses and thus have limited influence except making model fitting and convergence easier (87). In case of model convergence problems and model diagnostics (section 12.7) indicating that the problems are caused by covariates or nuisance parameters (which, e.g., may occur in early adaptive analyses for binary/categorical covariates with few participants having those values such as a *site* term for a newly started site that has only included few participants), neutral but more informative priors may be used for these model terms at a specific analysis (with clear description of this in the report) or these variables may be temporarily combined (e.g., sites with few participants may be combined) or omitted (e.g., adjustment variables with few values in a category). In these cases, the priors outlined below, and all variables will be used again as soon as possible.

Primary priors for the logistic regression models

All binary outcomes will be analysed using logistic regression models with the following priors, specified on the log odds scale.

- Intercept: *N(0, 2.5)* [denoting a normally distributed prior with mean 0 and SD 2.5], corresponding to an event probability for a participant allocated to piperacillin/tazobactam and with all covariates set to 0/no/the reference site; as such a participant will not exist in the trial (due to age being set at

0), this prior is not directly interpretable, but extremely vague. The prior is centred at a probability of 50% with 95% probability mass between 0.7% and 99.3% and corresponds to essentially no information (less information than a single participant).

- Treatment effect: *N(0, 0.5)*, corresponding to distributions on the odds ratio scale centred at 1.00 (no difference) and with 95% probability mass between 0.38 and 2.66. This prior corresponds to approximately 85 participants in a two-arm trial with fixed equal allocation and identical event probabilities of 25% in both arms analysed using a conventional logistic regression model (88, 137) (event probability 20%: 100 participants; 30%: 76 participants). This prior has been based on a simulation study with similarities to EMPRESS (75).
- Age (years) and age squared: a *N(0, 0.15)* for each 1-unit increase; these priors are very vague considering the expected substantial range of age and age squared and corresponds to odds ratio for each 1-unit change centred at 1.00 (no difference) and with 95% probability mass between 0.75 and 1.34.
- All binary and categorical adjustment variables: *N(0, 1)* priors, corresponding to distributions on the odds ratio scale centred at 1.00 and with 95% probability mass between 0.14 and 7.10.

Primary priors for the linear regression models:

- Intercept: *N(x, y)* priors with *x* corresponding to the midpoint of the range of possible values for each outcome, and *y* corresponding to the distance from the midpoint to either end of the range of possible values (with proportions used for binary outcomes). This prior corresponds to the prior mean value for a participant allocated to piperacillin/tazobactam and with all covariates set to 0/no/the reference site; as such a participant will not exist in the trial (due to age being set at 0), this prior is not directly interpretable, but extremely vague. The prior is centred at *x* (the midpoint for each scale) and the 68% central probability mass covers the range of each scale. The prior thus also includes implausible values, but as it corresponds to minimal information, it will quickly be overwhelmed by the data.

  For EQ-5D-5L index values, most values are between 0 and 1, although few patients usually have values <0 (50, 52, 57). For this outcome, we will define both *x* and *y* as 0.5 (i.e., using the 0-1 range of the scale and the strategy outlined above).
- Treatment effect: *N(0, z)* priors with *z* corresponding to the 15% of the range of the scale of each outcome (using 15% of 0 to 1 for EQ-5D-5L index values) corresponding to prior probability distributions for mean differences centred at 0 and with 95% probability mass between -29.4% and

29.4% of the full outcome range, e.g., for days alive out of hospital at day 90, 95% prior probability mass will be centred between -26.5 and 26.5 days.

- Age (years) and age squared: a *N(0, z)* for each 1 unit increase as for the treatment effect, with *z* corresponding to 2.5% of the range of the scale of each outcome (as outlined above, using 2.5% of 0 to 1 for EQ-5D-5L index values), corresponding to prior probability distributions for mean differences centred at 0 and with 95% probability mass between -4.41% and 4.41% of the full outcome range for each 1 unit increase; these priors are very vague considering to the expected substantial range of age and age squared.
- All binary and categorical adjustment variables: *N(0, z)* priors, with *z* defined as twice as large as the primary priors for the treatment effects, i.e., 30% of the range of scale of each outcome (using 30% of 0 to 1 for EQ-5D-5L index values) corresponding to prior probability distributions for mean differences centred at 0 and with 95% probability mass between -58.8% and 58.8% of the full outcome range, e.g., for days alive out of hospital at day 90, 95% prior probability mass will be centred between -52.9 and 52.9 days.
- Sigma: *half-N(y)* [a half-normal distribution, i.e., a folded normal distribution with mean 0/a normal distribution truncated to only contain non-negative values, with SD = y], with *y* defined as for the intercept. This prior conveys very minimal information and will have minimal influence on the results.

Sensitivity analyses

Several sensitivity analyses will be conducted using different priors for the primary, final analyses of each outcome, using different priors for the treatment effects but identical priors for all other model terms.

More sceptical priors

A set of sensitivity analyses using *more* sceptical priors for the treatment effects will be conducted. For the binary outcomes, *N(0, 0.15)* priors will be used, which corresponds to prior probability distributions on the odds ratio scale centred at 1.00 (no difference) and with 95% probability mass between 0.75 and 1.34; this corresponds to the same amount of information as 948 participants in a trial as described above with 25% event probabilities in both arms (20%: 1111 participants; 30%: 847 participants). For the continuous outcomes, *N(0, z)* priors will be used, with *z* corresponding to 2.5% of the range of the scale of each outcome (as outlined above, using 2.5% of 0 to 1 for EQ-5D-5L index values), corresponding to prior probability distributions for mean differences centred at 0 and with 95% probability mass between -4.9%

and 4.9% of the full outcome range, e.g., for days alive out of hospital at day 90, 95% prior probability mass will be centred between -4.4 and 4.4 days.

Less informative priors

A set of sensitivity analyses using *less* informative/sceptical priors, i.e., essentially non-informative priors for the treatment effects, will also be conducted. For the binary outcomes, *N(0, 2.5)* priors will be used, which corresponds to prior probability distributions on the odds ratio scale centred at 1.00 (no difference) and with 95% probability mass between < 0.01 and 134; this corresponds to the same amount of information as 3.4 participants in a trial as described above with 25% event probabilities in both arms (20%: 4 participants; 30%: 3 participants). For the continuous outcomes, *N(0, z)* priors will be used, with *z* corresponding to 50% of the range of the scale of each outcome (as outlined above, using 50% of 0 to 1 for EQ-5D-5L index values) corresponding to prior probability distributions for mean differences centred at 0 and with 95% probability mass between -98% and 98% of the full outcome range, e.g., for days alive out of hospital at day 90, 95% prior probability mass will be centred between -88 and 88 days.

Evidence-based priors

We will consider a set of sensitivity analyses using *evidence-based priors*, i.e., priors based on summarised data from previous trials based on, e.g., an updated meta-analysis. At the time of writing, the most recent meta-analyses available have been conducted by our group (22), and includes data on some of the clinical outcomes assessed in EMPRESS. While it included relevant trials, they were generally conducted in patients who were ill enough to require hospital admission but with less severe disease than participants in EMPRESS, and thus the evidence is somewhat indirect. When EMPRESS concludes, we expect that additional relevant trials may have emerged based on several identified ongoing trials in the aforementioned systematic review with meta-analyses (22). Consequently, we have opted to not specify exact priors for any sensitivity analyses using evidence-based priors at this time, but will define these later if relevant.

Priors for analyses of heterogeneity in treatment effects

In addition to the analyses assessing average treatment effects, additional analyses assessing heterogeneity in treatment effects (HTE) according to the baseline covariates outlined in section 12.8, the following additional priors will be used (for all other model terms, the priors outlined above will be used).

Priors for the HTE analyses according to continuous baseline variables:

For all additional model terms (the main term of the baseline covariate and its quadratic transformation) and the interaction terms, *N(0, 0.15)* priors will be used for each 1-unit increase (after log2-transformation of plasma creatinine); as such, these priors are very weak and correspond to the priors for age and age squared specified above. These priors will have minimal influence on the results and be quickly overwhelmed by the data.

Priors for the HTE analyses according to categorical baseline variables:

For the subgroup-level intercepts and treatment effects *N(0, omega)* priors will be used, with *omega* being the shrinkage factor estimated from the data and using a *half-N(0.5)* hyper-prior; this corresponds to a SD (omega) that with 95% probability is between 0.02 and 1.12 with mean 0.40 and median 0.34. This prior will similarly have relatively limited influence on the results given the amount of data.

19.9 Appendix 9: Simulation-based assessment of trial design performance

The EMPRESS trial design was developed and evaluated using statistical simulation as recommended (25, 66, 67, 80, 138). In this appendix, details regarding the simulations are presented along with the final trial design (and several additional variants which were also evaluated), and the final performance metrics.
All statistical simulations were conducted in R v4.2.2 (R Core Team, R Foundation for Statistical Computing, Vienna, Austria) using the *adaptr* package (inceptdk.github.io/adaptr) v1.3.0 (80) developed by members of the EMPRESS management committee. For all trial designs evaluated, we conducted 100,000 simulations as recommended (67). The design choices and assumptions are outlined for the final trial design below with different values or design choices detailed for each alternative design presented.

Model and prior

All simulations used *Beta-binomial* conjugate priors (107) to facilitate simulation speed and as simulation of/adjustment for baseline covariates is omitted (as is general practice) during the simulations; 20,000 independent posterior samples were drawn for each trial arm during each analysis and used for all comparisons and adaptation rules (80). The priors were specified to correspond to the same amount of information as the priors that will be used in the primary analyses of the trial (*N(0, 0.5)*, i.e., a normally distributed prior with mean 0 and SD 0.5) on the log-odds scale in the logistic regression model planned), and thus conveying the same amount of scepticism while being neutral. The planned prior corresponds to the same amount of information as a two-arm trial with 1:1 allocation and 25% event probabilities in both groups including approximately 85 participants in total. Thus, the prior in each arm was specified by calculating the observed event probability across both arms at the time of each analysis, which was used to calculate the number of participants that, given this event probability, would correspond to the desired prior (137). This estimated number of participants was divided by two (half in each arm), which was in turn multiplied by the observed event probability and the observed non-event probability, to set the *alpha* and *beta* parameters of the *Beta* prior distribution accordingly. Consequently, the prior used in each analysis corresponds to the intended level of information, conveys some scepticism pulling estimates in the two arms closer to the overall event probability and thus each other.

Timing of analyses

The first adaptive analyses were conducted once outcome data were available for 400 simulated participants, with subsequent analyses after each 300 participants until a maximum of 14,000 participants. The number of participants *randomised* at the time of each adaptive analyses was calculated assuming a total outcome-data lag of 45 days (follow-up duration of 30 days plus 15 days of data collection/validation)

and a constant inclusion rate of 5 participants per day, with the rate and assumption of this being constant (following initiation of sites) based on previous trials by our group (33-37). The longer the outcome-data lag and the higher the inclusion rate, the larger the proportion of participants randomised (using the allocation profile active at the time of their randomisation) but without outcome data available at the time of an adaptive analysis, with larger proportions increasing the potential for subsequent important changes in the estimates once all randomised participants are later included in a final analysis (85).

Scenarios

The performance of the primary EMPRESS trial design was evaluated under three clinical scenarios as described in section 12.6:

- No difference: 25% event probability in both groups
- Small difference (corresponding to the threshold for practical equivalence outlined in section 12.4): 25% event probability in the piperacillin/tazobactam group, 22.5% event probability in the meropenem group (10% relative risk reduction)
- Large difference: 25% event probability in the meropenem group, 20% event probability in the meropenem group (20% relative risk reduction)

Adaptation rules

We used constant, symmetrical stopping rules for superiority and inferiority with the probability thresholds for stopping *calibrated* to obtain a probability of stopping for superiority of 5% in the scenario without differences (the Bayesian analogue of the type 1 error rate) (25). Calibration was performed using a Gaussian process-based Bayesian optimisation algorithm (139) with 100,000 simulations during each calibration step and an accepted final tolerance range of 4.9-5.0% probability of superiority. The calibrated stopping thresholds for superiority are displayed with the performance metrics below (rounded to a maximum of 6 decimals) with the calibrated stopping thresholds for inferiority defined as 1 minus the threshold for superiority.

Trials were stopped for practical equivalence if there was >90% probability of an absolute difference less than 2.5%-points at any analysis unless the threshold for superiority was declared at that analysis.

The initial allocation profile consisted of 50% : 50% allocation. Response-adaptive randomisation restricted to minimum 40% allocation to each arm (without additional softening) were be used from the first adaptive analysis and onwards based on the posterior probabilities of each arm being the overall best (25).

Performance metrics

Multiple performance metrics were calculated as previously described (25, 80) based on 100,000 simulations each and presented for the assessed designs. In brief, the following metrics are presented:

1) Sample size (mean [expected], SD, median, 25th percentile [P25], 75th percentile [P75], minimum, maximum)
2) Event counts (total event counts in both arms in the simulated trials; same summary measures as for #1)
3) Event probabilities (total event probabilities across both arms in the simulated trials, i.e., event counts/sample size; same summary measures as for #1)
4) Pr(conclusive): the probability of conclusiveness, i.e., the proportion of simulated trials triggering the superiority or practical equivalence rule at any adaptive analysis.
5) Pr(superiority): the probability of superiority, i.e., the proportion of simulated trials triggering the superiority stopping rule at any adaptive analysis. This corresponds to the Bayesian analogues of the type 1 error rate in scenarios with no difference present and the power in scenarios with a difference present (25).
6) Pr(equivalence): the probability of practical equivalence, i.e., the proportion of simulated trials triggering the stopping rule for practical equivalence at any adaptive analysis.
7) Pr(max): the proportion of simulated trials reaching the maximum allowed sample size without triggering a stopping rule, i.e., the probability of the trial being inconclusive.
8) Pr(piperacillin/tazobactam superior): the probability of declaring piperacillin/tazobactam superior, i.e., the proportion of simulated trials with the piperacillin/tazobactam arm triggering the stopping rule for superiority.
9) Pr(meropenem superior): the probability of declaring meropenem superior, i.e., the proportion of simulated trials with the meropenem arm triggering the stopping rule for superiority.
10) Pr(none superior): the probability of not declaring any arm superior, i.e., the proportion of simulated trials stopped due to triggering the equivalence stopping rule or reaching the maximum sample size without triggering any stopping rule.
11) RMSE (superiority only): the root mean squared error (RMSE) of the estimated versus the true simulated event probability in the superior arm (only calculated for trials stopped for superiority; %-points).
12) RMSE (select piperacillin/tazobactam in inconclusive trials): the root mean squared error (RMSE) of the estimated versus the true simulated event probability in the selected arm, with the piperacillin/tazobactam (control or standard of care) selected if not stopped for superiority (%-points).

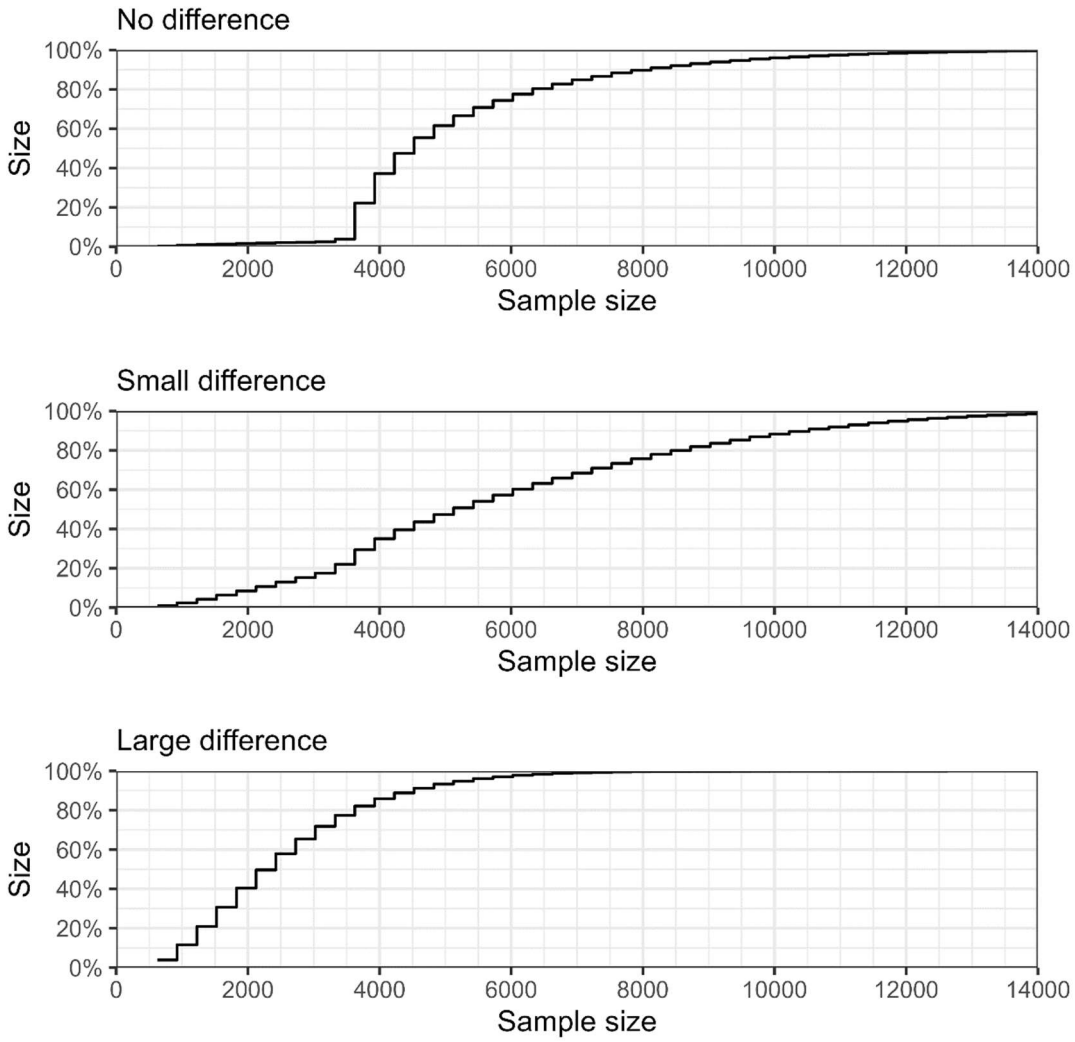
13) RMSE (treatment effect): the RMSE of the estimated difference with meropenem compared to piperacillin/tazobactam versus the true simulated difference (for trials ending in superiority of meropenem only; %-points).
14) IDP (superiority only): the ideal design percentage (IDP, %), calculated as previously described (25, 86, 140) for trials ending in superiority only. In brief, the IDP is 100% for a design that always ends selecting the best arm and lower otherwise and can thus only be calculated for scenarios with differences between arms. For trials with >2 arms, selecting arms that are not the best will lead to decreases proportional to how much worse the selected arm is (i.e., selecting a slightly inferior arm will decrease IDP less than selecting a substantially inferior arm). This metric is less important in two-arm trials like EMPRESS but supplements the other metrics.
15) IDP (select piperacillin/tazobactam in inconclusive trials): the IDP calculated for all trials, considering the piperacillin/tazobactam (control or standard of care) arm selected in trials not stopped for superiority.

Variants

Several variants of the final EMPRESS design (**Table S4**) were evaluated using simulation; results are briefly summarised here.

First, the final trial design was re-evaluated with the calibrated stopping thresholds for superiority and inferiority rounded to 4 decimals places (as the calibration process itself puts no limit on the number of decimals), with essentially unchanged results, including the type 1 error rate and power (**Table S5**). Consequently, these rounded thresholds will be used in the actual trial. The empirical cumulative distribution functions for the sample sizes with this design under all three clinical scenarios are presented in **Figure S2** below.

Figure S2: Empirical cumulative distribution functions for sample sizes



Empirical cumulative distribution functions for sample sizes in the primary trial design with rounded stopping rules under all three clinical scenarios. The values on the vertical axes correspond to the percentage of simulated trials ending with a sample size equal to or lower than the value on the horizontal axes.

Second, the final trial design and calibrated stopping rules were used after changing the event probability in the piperacillin/tazobactam group to first 20% (**Table S6**) and then 30% (**Table S7**) with 10% relative risk reductions with meropenem used in the small-difference scenario and as the equivalence difference (event probabilities 18% and 27%, respectively; equivalence differences of 2% and 3%, respectively) and 20% relative risk reductions in the large-difference scenario (event probabilities of 16% and 24%, respectively). This substantially affected the sample sizes and overall event probabilities with small effects on the total

event counts, and, importantly, negligible effects on the type 1 error rates and acceptable effects on the power.

Third, we evaluated variants of the final EMPRESS design using first fixed, equal randomisation (**Table S8**) and less restricted response-adaptive randomisation, i.e., minimum allocation of 35%, but a softening factor of 0.7 (25) (**Table S9**). The fixed design led to slightly lower expected sample sizes in all scenarios (approximately 100 participants in the no-difference scenario and <20 participants in the large-difference scenario), while expected total event counts where slightly but negligibly different (approximately 25 and 8 participants lower in the no- and small-difference scenarios and 6 participants higher in the large-difference scenario); expected total event probabilities were slightly higher with the fixed design in both scenarios with differences present. The design with less restricted response-adaptive randomisation led to slightly higher expected sample sizes in all scenarios (30-100 participants less, depending on the scenario) and slightly higher expected event counts in all scenarios (approximately 3 to 26 events); expected total event probabilities were slightly lower in both scenarios with differences present. Other performance metrics were largely unaffected. Overall, using fixed randomisation or less restricted response-adaptive randomisation had no major influence on the performance metrics; based on an overall consideration about ethical and logistical considerations and the potential influence of less restricted response-adaptive randomisation on clinical equipoise (i.e., with less restricted response-adaptive randomisation, it may be easier to guess which treatment arm currently has the highest probability of being superior given the unblinded design), the final design uses relatively restricted response-adaptive randomisation to appropriately balance all considerations.

Fourth, we evaluated a variant of the final EMPRESS trial design with analyses conducted after the first 400 participants and then after each additional 600 participants (**Table S10**), to assess the influence on fewer analyses (and thus less strict stopping thresholds to maintain the type 1 error rate at 5%). This slightly increased the expected sample sizes and total event counts despite the slightly less stringent stopping thresholds.

Fifth, we evaluated a variant of the final EMPRESS trial design with a higher inclusion rate of 10 participants/day (**Table S11**), meaning that the proportion of randomised participants with outcome data available at the time of analysis would be lower. This slightly increased the mean sample sizes and mean total event counts and led to small but unimportant deteriorations in some of the other performance metrics.

Sixth, we evaluated a variant of the final EMPRESS trial design with a maximum sample size of 10,000 participants (**Table S12**), and, thus, also a lower maximum possible number of adaptive analyses. This slightly decreased the expected sample sizes and total event counts in all scenarios, but at the cost of

increasing the proportion of inconclusive trials to 3.3% in the scenario without differences and 9.7% in the scenario with small differences, which we considered too high.

Seventh, we evaluated a variant of the final EMPRESS trial design with a stricter probability threshold of >95% for practical equivalence (**Table S13**). This substantially increased the expected sample sizes and total event counts in the no- and small-difference scenarios due to lowering the probabilities of stopping for practical equivalence from ~95% to ~93% in the no-difference scenario and from ~27% to ~14% in the small-difference scenario with the stops occurring later. In addition, the probability of inconclusiveness increased to ~2% and ~5% in the no- and small difference scenarios. Consequently, this stricter threshold was omitted.

Finally, we evaluated a variant of the final EMPRESS trial design with a total of 105 days outcome-data lag, corresponding to the use of 90-day mortality as the guiding outcome (assuming the same event probabilities as in the primary design) and 15 days of data collection/verification lag (**Table S14**). This increased the expected sample sizes with approximately 300 participants in each scenario and the expected total event counts with approximately 65-70 in each scenario; total event probabilities and probabilities of different conclusions were similar. As we expect a relatively short intervention period, we expect most intervention-related deaths to occur relatively close to the time of randomisation (i.e., within 30 days). Moreover, as using 90-day mortality increased total sample sizes and event counts, we decided to ultimately use 30-day mortality as the primary and guiding outcome.

Table S4: Performance metrics of the final EMPRESS trial design

Metric	No difference	Small difference	Large difference
Sample size – mean	5190.8	5858.8	2570.4
Sample size – SD	2069.5	3127.5	1417.9
Sample size – median	4525	5125	2425
Sample size – P25	3925	3625	1525
Sample size – P75	6025	7825	3325
Sample size – minimum	625	625	625
Sample size – maximum	14000	14000	12625
Event count – mean	1297.7	1380.4	569.3
Event count – SD	519.3	736.3	312.6
Event count – median	1120	1227.5	507
Event count – P25	961	852	336
Event count – P75	1486	1855	739
Event count – minimum	123	115	108
Event count – maximum	3673	3444	2792
Event probability – mean	0.25	0.236	0.222
Event probability – SD	0.00647	0.00681	0.00968
Event probability – median	0.25	0.236	0.222
Event probability – P25	0.246	0.232	0.216
Event probability – P75	0.254	0.24	0.228
Event probability – minimum	0.197	0.184	0.173
Event probability – maximum	0.296	0.294	0.291
Pr(conclusive)	0.997	0.99	1
Pr(superiority)	0.0495	0.723	0.998
Pr(equivalence)	0.948	0.267	0.00222
Pr(max)	0.00282	0.00962	0
Pr(piperacillin/tazobactam superior)	0.0247	0.00048	0.00001
Pr(meropenem superior)	0.0249	0.723	0.998
Pr(none superior)	0.95	0.277	0.00222
RMSE (superiority only)	0.0235	0.011	0.0123
RMSE (select piperacillin/tazobactam in inconclusive trials)	0.023	0.0109	0.0123
RMSE (treatment effect)	0.0484	0.0191	0.0186
IDP (superiority only)	-	99.9	100
IDP (select piperacillin/tazobactam in inconclusive trials)	-	98.6	100

Calibrated stopping threshold for superiority: 0.996409.

Table S5: Performance metrics of the EMPRESS design (stopping rules rounded to 4 decimals)

Metric	No difference	Small difference	Large difference
Sample size – mean	5188.8	5858.8	2570.4
Sample size – SD	2067.5	3127.6	1417.9
Sample size – median	4525	5125	2425
Sample size – P25	3925	3625	1525
Sample size – P75	6025	7825	3325
Sample size – minimum	625	625	625
Sample size – maximum	14000	14000	12625
Event count – mean	1297.3	1380.4	569.3
Event count – SD	518.8	736.3	312.6
Event count – median	1120	1227	507
Event count – P25	961	852	336
Event count – P75	1485.2	1855	739
Event count – minimum	123	115	108
Event count – maximum	3673	3444	2792
Event probability – mean	0.25	0.236	0.222
Event probability – SD	0.00647	0.00681	0.00968
Event probability – median	0.25	0.236	0.222
Event probability – P25	0.246	0.232	0.216
Event probability – P75	0.254	0.24	0.228
Event probability – minimum	0.197	0.184	0.173
Event probability – maximum	0.296	0.294	0.291
Pr(conclusive)	0.997	0.99	1
Pr(superiority)	0.0499	0.723	0.998
Pr(equivalence)	0.947	0.267	0.00222
Pr(max)	0.00278	0.00962	0
Pr(piperacillin/tazobactam superior)	0.025	0.00049	0.00001
Pr(meropenem superior)	0.0249	0.723	0.998
Pr(none superior)	0.95	0.277	0.00222
RMSE (superiority only)	0.0235	0.011	0.0123
RMSE (select piperacillin/tazobactam in inconclusive trials)	0.023	0.0109	0.0123
RMSE (treatment effect)	0.0484	0.0191	0.0186
IDP (superiority only)	-	99.9	100
IDP (select piperacillin/tazobactam in inconclusive trials)	-	98.6	100

The primary EMPRESS design including calibrated stopping thresholds, but as the calibration process does not include a limit on the number of decimals, the calibrated stopping thresholds for superiority and inferiority were rounded to 4 decimal places (as will be used during the actual analyses), to assess that this has minimal influence on performance metrics. Calibrated stopping threshold for superiority: 0.9964.

Table S6: Performance metrics of the EMPRESS design assuming lower event probabilities

Metric	No difference	Small difference	Large difference
Sample size – mean	6669.6	7315.3	3228.8
Sample size – SD	2532.1	3722.6	1862.9
Sample size – median	5725	6625	2725
Sample size – P25	4825	4525	1825
Sample size – P75	7825	10225	4225
Sample size – minimum	625	625	625
Sample size – maximum	14000	14000	14000
Event count – mean	1334	1378.9	571.5
Event count – SD	509	701.4	328.6
Event count – median	1160	1251	502
Event count – P25	995	864	324
Event count – P75	1539	1906	749
Event count – minimum	95	91	84
Event count – maximum	2975	2820	2548
Event probability – mean	0.2	0.189	0.177
Event probability – SD	0.00531	0.00565	0.00812
Event probability – median	0.2	0.189	0.177
Event probability – P25	0.197	0.185	0.172
Event probability – P75	0.203	0.192	0.182
Event probability – minimum	0.152	0.141	0.134
Event probability – maximum	0.248	0.232	0.237
Pr(conclusive)	0.976	0.926	1
Pr(superiority)	0.052	0.672	0.997
Pr(equivalence)	0.924	0.254	0.00253
Pr(max)	0.0244	0.0744	0.00004
Pr(piperacillin/tazobactam superior)	0.0261	0.00071	0.00005
Pr(meropenem superior)	0.0259	0.671	0.997
Pr(none superior)	0.948	0.328	0.00257
RMSE (superiority only)	0.0209	0.00966	0.0104
RMSE (select piperacillin/tazobactam in inconclusive trials)	0.0177	0.00926	0.0104
RMSE (treatment effect)	0.0425	0.0171	0.0161
IDP (superiority only)	-	99.9	100
IDP (select piperacillin/tazobactam in inconclusive trials)	-	89.8	100

The primary EMPRESS design using the primary calibrated stopping thresholds for superiority and inferiority, but with a 20% simulated event probability in the piperacillin/tazobactam group in all scenarios, with simulated event probabilities in the meropenem group of 20% (no difference scenario), 18% (small difference scenario, 10% relative risk reduction), and 16% (large difference scenario, 20% relative risk reduction). Practical equivalence defined as 2%-points, i.e., a 10% relative risk reduction. Calibrated stopping threshold for superiority: 0.996409 (not re-calibrated).

Table S7: Performance metrics of the EMPRESS design assuming higher event probabilities

Metric	No difference	Small difference	Large difference
Sample size – mean	4156.8	4740.5	2131.2
Sample size – SD	1649.4	2501	1123.9
Sample size – median	3625	4225	1825
Sample size – P25	3025	3025	1225
Sample size – P75	4825	6325	2725
Sample size – minimum	625	625	625
Sample size – maximum	14000	14000	10825
Event count – mean	1247.2	1340.5	567
Event count – SD	496.7	706.3	297.4
Event count – median	1081	1195	500
Event count – P25	920	822	336
Event count – P75	1423	1789	727
Event count – minimum	161	143	129
Event count – maximum	4291	4090	2879
Event probability – mean	0.3	0.283	0.267
Event probability – SD	0.00756	0.00786	0.0111
Event probability – median	0.3	0.283	0.266
Event probability – P25	0.295	0.278	0.26
Event probability – P75	0.305	0.288	0.273
Event probability – minimum	0.258	0.229	0.206
Event probability – maximum	0.357	0.342	0.333
Pr(conclusive)	1	1	1
Pr(superiority)	0.0463	0.73	0.998
Pr(equivalence)	0.954	0.27	0.00219
Pr(max)	0.00002	0.00001	0
Pr(piperacillin/tazobactam superior)	0.023	0.00023	0
Pr(meropenem superior)	0.0233	0.73	0.998
Pr(none superior)	0.954	0.27	0.00219
RMSE (superiority only)	0.0264	0.0123	0.0141
RMSE (select piperacillin/tazobactam in inconclusive trials)	0.0263	0.0123	0.0141
RMSE (treatment effect)	0.0531	0.021	0.0207
IDP (superiority only)	-	100	100
IDP (select piperacillin/tazobactam in inconclusive trials)	-	100	100

The primary EMPRESS design using the primary calibrated stopping thresholds for superiority and inferiority, but with a 30% simulated event probability in the piperacillin/tazobactam group in all scenarios, with simulated event probabilities in the meropenem group of 30% (no difference scenario), 27% (small difference scenario, 10% relative risk reduction), and 24% (large difference scenario, 20% relative risk reduction). Practical equivalence defined as 3%-points, i.e., a 10% relative risk reduction. Calibrated stopping threshold for superiority: 0.996409 (not re-calibrated).

Table S8: Performance metrics of the EMPRESS if using fixed, equal allocation profile

Metric	No difference	Small difference	Large difference
Sample size – mean	5092.9	5779	2557.8
Sample size – SD	1983.7	3065.7	1392.4
Sample size – median	4525	5125	2425
Sample size – P25	3925	3625	1525
Sample size – P75	5725	7825	3325
Sample size – minimum	625	625	625
Sample size – maximum	14000	14000	11725
Event count – mean	1273.2	1372.5	575.5
Event count – SD	498.1	729.3	313.3
Event count – median	1103	1224	512
Event count – P25	947	846	341
Event count – P75	1451	1837	748
Event count – minimum	117	118	103
Event count – maximum	3658	3513	2554
Event probability – mean	0.25	0.237	0.225
Event probability – SD	0.00652	0.00676	0.00962
Event probability – median	0.25	0.238	0.225
Event probability – P25	0.246	0.233	0.219
Event probability – P75	0.254	0.241	0.231
Event probability – minimum	0.187	0.182	0.165
Event probability – maximum	0.296	0.299	0.286
Pr(conclusive)	0.998	0.993	1
Pr(superiority)	0.0496	0.724	0.998
Pr(equivalence)	0.949	0.269	0.00206
Pr(max)	0.00182	0.00669	0
Pr(piperacillin/tazobactam superior)	0.0248	0.00045	0.00002
Pr(meropenem superior)	0.0248	0.724	0.998
Pr(none superior)	0.95	0.276	0.00206
RMSE (superiority only)	0.0256	0.0116	0.0127
RMSE (select piperacillin/tazobactam in inconclusive trials)	0.0252	0.0115	0.0127
RMSE (treatment effect)	0.0473	0.0189	0.0185
IDP (superiority only)	-	99.9	100
IDP (select piperacillin/tazobactam in inconclusive trials)	-	99	100

Variant of the final EMPRESS design using simple randomisation with a fixed, equal allocation profile, i.e., 50%:50% allocation. Calibrated stopping threshold for superiority: 0.996581 (re-calibrated).

Table S9: Performance metrics of the EMPRESS if using less restricted response-adaptive randomisation

Metric	No difference	Small difference	Large difference
Sample size – mean	5291.3	5964.7	2605.1
Sample size – SD	2163.6	3216.7	1467
Sample size – median	4525	5425	2425
Sample size – P25	3925	3625	1525
Sample size – P75	6025	8125	3325
Sample size – minimum	625	625	625
Sample size – maximum	14000	14000	13825
Event count – mean	1323	1400	572.5
Event count – SD	542.9	753.1	320.3
Event count – median	1135	1245	507
Event count – P25	970	856	333
Event count – P75	1519	1889	746
Event count – minimum	127	111	99
Event count – maximum	3662	3438	3045
Event probability – mean	0.25	0.235	0.221
Event probability – SD	0.00647	0.00681	0.00973
Event probability – median	0.25	0.235	0.22
Event probability – P25	0.246	0.231	0.214
Event probability – P75	0.254	0.239	0.226
Event probability – minimum	0.203	0.178	0.158
Event probability – maximum	0.301	0.301	0.283
Pr(conclusive)	0.996	0.986	1
Pr(superiority)	0.0498	0.72	0.998
Pr(equivalence)	0.946	0.266	0.00234
Pr(max)	0.00435	0.0139	0
Pr(piperacillin/tazobactam superior)	0.0246	0.00041	0
Pr(meropenem superior)	0.0253	0.72	0.998
Pr(none superior)	0.95	0.28	0.00234
RMSE (superiority only)	0.0227	0.0107	0.0122
RMSE (select piperacillin/tazobactam in inconclusive trials)	0.0218	0.0106	0.0122
RMSE (treatment effect)	0.0484	0.0193	0.0187
IDP (superiority only)	-	99.9	100
IDP (select piperacillin/tazobactam in inconclusive trials)	-	98	100

Variant of the final EMPRESS design using less restricted response-adaptive randomisation, with minimum 35% allocation and restricted by using a softening factor of 0.7.
Calibrated stopping threshold for superiority: 0.996257 (re-calibrated).

Table S10: Performance metrics of the EMPRESS if conducting analyses after every 600 participants

Metric	No difference	Small difference	Large difference
Sample size – mean	5435.7	6075.8	2698
Sample size – SD	2149.8	3136.5	1416
Sample size – median	4825	5425	2425
Sample size – P25	4225	3625	1825
Sample size – P75	6025	8425	3625
Sample size – minimum	625	625	625
Sample size – maximum	14000	14000	12025
Event count – mean	1359	1431.7	597.6
Event count – SD	539.5	738.1	312.4
Event count – median	1197	1288	538
Event count – P25	1026	868	391
Event count – P75	1555	1918	784
Event count – minimum	119	113	101
Event count – maximum	3648	3475	2647
Event probability – mean	0.25	0.236	0.222
Event probability – SD	0.00629	0.0066	0.00938
Event probability – median	0.25	0.236	0.222
Event probability – P25	0.246	0.232	0.216
Event probability – P75	0.254	0.24	0.228
Event probability – minimum	0.19	0.181	0.162
Event probability – maximum	0.288	0.293	0.288
Pr(conclusive)	0.997	0.991	1
Pr(superiority)	0.0499	0.737	0.998
Pr(equivalence)	0.947	0.254	0.00188
Pr(max)	0.00338	0.00939	0
Pr(piperacillin/tazobactam superior)	0.0244	0.00042	0
Pr(meropenem superior)	0.0254	0.736	0.998
Pr(none superior)	0.95	0.263	0.00188
RMSE (superiority only)	0.0224	0.0104	0.0119
RMSE (select piperacillin/tazobactam in inconclusive trials)	0.0216	0.0104	0.0119
RMSE (treatment effect)	0.0453	0.018	0.0179
IDP (superiority only)	-	99.9	100
IDP (select piperacillin/tazobactam in inconclusive trials)	-	98.5	100

Variant of the final EMPRESS design with adaptive analyses starting after the first 400 participants have available data (as in the main design), followed by adaptive analyses after every 600 additional participants.
Calibrated stopping threshold for superiority: 0.995661 (re-calibrated).

Table S11: Performance metrics of the EMPRESS assuming 10 participants included per day

Metric	No difference	Small difference	Large difference
Sample size – mean	5404.5	6103.2	2801.1
Sample size – SD	2077.8	3122.7	1417.5
Sample size – median	4750	5350	2650
Sample size – P25	4150	3850	1750
Sample size – P75	6250	8050	3550
Sample size – minimum	850	850	850
Sample size – maximum	14000	14000	12550
Event count – mean	1351.1	1438.4	621.3
Event count – SD	521.7	735	312.5
Event count – median	1173	1286	558
Event count – P25	1013	908	388
Event count – P75	1534	1912	789
Event count – minimum	183	159	143
Event count – maximum	3657	3459	2774
Event probability – mean	0.25	0.236	0.222
Event probability – SD	0.00629	0.00643	0.00899
Event probability – median	0.25	0.236	0.222
Event probability – P25	0.246	0.232	0.217
Event probability – P75	0.254	0.24	0.228
Event probability – minimum	0.214	0.187	0.168
Event probability – maximum	0.295	0.284	0.282
Pr(conclusive)	0.997	0.99	1
Pr(superiority)	0.0497	0.719	0.998
Pr(equivalence)	0.947	0.271	0.00242
Pr(max)	0.00295	0.00966	0
Pr(piperacillin/tazobactam superior)	0.0247	0.00046	0.00002
Pr(meropenem superior)	0.025	0.719	0.998
Pr(none superior)	0.95	0.281	0.00242
RMSE (superiority only)	0.021	0.00998	0.0113
RMSE (select piperacillin/tazobactam in inconclusive trials)	0.0205	0.00993	0.0113
RMSE (treatment effect)	0.0423	0.0169	0.0167
IDP (superiority only)	-	99.9	100
IDP (select piperacillin/tazobactam in inconclusive trials)	-	98.6	100

Variant of the final EMPRESS design assuming that 10 participants will be included per day, meaning that the proportion of randomised participants with available data at each adaptive analysis is lower.
Calibrated stopping threshold for superiority: 0.996549 (re-calibrated).

Table S12: Performance metrics of the EMPRESS with a maximum sample size of 10,000 participants

Metric	No difference	Small difference	Large difference
Sample size – mean	5110.9	5580.1	2540.4
Sample size – SD	1851.6	2693.1	1407
Sample size – median	4525	5125	2125
Sample size – P25	3925	3625	1525
Sample size – P75	6025	7825	3325
Sample size – minimum	625	625	625
Sample size – maximum	10000	10000	10000
Event count – mean	1277.9	1314.9	562.7
Event count – SD	465.2	634.1	310.2
Event count – median	1120	1214	496
Event count – P25	960	846	333
Event count – P75	1482	1828	733
Event count – minimum	125	111	107
Event count – maximum	2666	2514	2320
Event probability – mean	0.25	0.236	0.222
Event probability – SD	0.00651	0.00689	0.00972
Event probability – median	0.25	0.236	0.222
Event probability – P25	0.246	0.232	0.216
Event probability – P75	0.254	0.24	0.228
Event probability – minimum	0.2	0.178	0.171
Event probability – maximum	0.304	0.291	0.302
Pr(conclusive)	0.967	0.903	1
Pr(superiority)	0.0497	0.656	0.998
Pr(equivalence)	0.917	0.247	0.00205
Pr(max)	0.0334	0.0974	0.00014
Pr(piperacillin/tazobactam superior)	0.0245	0.00045	0
Pr(meropenem superior)	0.0252	0.655	0.998
Pr(none superior)	0.95	0.344	0.00219
RMSE (superiority only)	0.0241	0.0115	0.0124
RMSE (select piperacillin/tazobactam in inconclusive trials)	0.0192	0.0109	0.0124
RMSE (treatment effect)	0.0495	0.02	0.0187
IDP (superiority only)	-	99.9	100
IDP (select piperacillin/tazobactam in inconclusive trials)	-	86.7	100

Variant of the final EMPRESS design with a maximum sample size of 10,000 participants meaning that the maximum number of adaptive analyses is lower, leading to slightly less restrictive stopping rules to maintain the same type 1 error rate.
Calibrated stopping threshold for superiority: 0.996186 (re-calibrated).

Table S13: Performance metrics of the EMPRESS with a stricter equivalence stopping rule

Metric	No difference	Small difference	Large difference
Sample size – mean	6935.5	6975.8	2620.8
Sample size – SD	2456.6	3685.3	1448.6
Sample size – median	6025	6325	2425
Sample size – P25	5125	4225	1525
Sample size – P75	7825	9625	3325
Sample size – minimum	625	625	625
Sample size – maximum	14000	14000	12625
Event count – mean	1733.9	1643.3	580.5
Event count – SD	616.3	867.5	319.4
Event count – median	1533	1517	519
Event count – P25	1322	990	340
Event count – P75	1995	2269	750
Event count – minimum	127	120	110
Event count – maximum	3681	3473	2784
Event probability – mean	0.25	0.236	0.222
Event probability – SD	0.00561	0.00641	0.00962
Event probability – median	0.25	0.236	0.222
Event probability – P25	0.246	0.232	0.216
Event probability – P75	0.254	0.239	0.228
Event probability – minimum	0.2	0.192	0.17
Event probability – maximum	0.293	0.291	0.288
Pr(conclusive)	0.978	0.947	1
Pr(superiority)	0.0493	0.802	1
Pr(equivalence)	0.928	0.145	0.00026
Pr(max)	0.0223	0.053	0
Pr(piperacillin/tazobactam superior)	0.0245	0.00047	0.00001
Pr(meropenem superior)	0.0248	0.801	1
Pr(none superior)	0.951	0.198	0.00026
RMSE (superiority only)	0.023	0.0104	0.0122
RMSE (select piperacillin/tazobactam in inconclusive trials)	0.0196	0.0102	0.0122
RMSE (treatment effect)	0.0465	0.0179	0.0185
IDP (superiority only)	-	99.9	100
IDP (select piperacillin/tazobactam in inconclusive trials)	-	93.7	100

Variant of the final EMPRESS design with a stricter probability threshold for stopping for equivalence, i.e., >95% probability of an absolute difference <2.5%-points required.
Calibrated stopping threshold for superiority: 0.996679 (re-calibrated).

Table S14: Performance metrics of the EMPRESS with 90-day mortality

Metric	No difference	Small difference	Large difference
Sample size – mean	5463.9	6138.3	2866.9
Sample size – SD	2048.2	3117.2	1420.8
Sample size – median	4825	5425	2425
Sample size – P25	4225	3925	1825
Sample size – P75	6325	8125	3625
Sample size – minimum	925	925	925
Sample size – maximum	14000	14000	13825
Event count – mean	1366.1	1446.9	636.1
Event count – SD	514.2	733.7	313.3
Event count – median	1191	1292	570
Event count – P25	1031	921	403
Event count – P75	1552	1915	804
Event count – minimum	187	176	159
Event count – maximum	3678	3465	3030
Event probability – mean	0.25	0.236	0.222
Event probability – SD	0.00623	0.00644	0.00883
Event probability – median	0.25	0.236	0.222
Event probability – P25	0.246	0.232	0.217
Event probability – P75	0.254	0.24	0.228
Event probability – minimum	0.202	0.19	0.172
Event probability – maximum	0.294	0.302	0.277
Pr(conclusive)	0.997	0.991	1
Pr(superiority)	0.05	0.721	0.998
Pr(equivalence)	0.947	0.27	0.0022
Pr(max)	0.00272	0.009	0
Pr(piperacillin/tazobactam superior)	0.0247	0.00055	0.00004
Pr(meropenem superior)	0.0253	0.72	0.998
Pr(none superior)	0.95	0.279	0.0022
RMSE (superiority only)	0.0205	0.00982	0.0111
RMSE (select piperacillin/tazobactam in inconclusive trials)	0.02	0.00978	0.0111
RMSE (treatment effect)	0.0412	0.0165	0.0163
IDP (superiority only)	-	99.9	100
IDP (select piperacillin/tazobactam in inconclusive trials)	-	98.6	100

Variant of the final EMPRESS design using a total of 105 days outcome-data lag, corresponding to 90-day mortality (same event probabilities as in the primary design) and 15 days of data collection lag.
Calibrated stopping threshold for superiority: 0.996367 (re-calibrated).