REVIEW

Randomised clinical trials in critical care: past, present and future

Anders Granholm[1]* , Waleed Alhazzani[2,3], Lennie P. G. Derde[4,5], Derek C. Angus[6], Fernando G. Zampieri[7],
Naomi E. Hammond[8,9], Rob Mac Sweeney[10], Sheila N. Myatra[11], Elie Azoulay[12], Kathryn Rowan[13],
Paul J. Young[14,15,16,17], Anders Perner[1] and Morten Hylander Møller[1]

Abstract

Randomised clinical trials (RCTs) are the gold standard for providing unbiased evidence of intervention effects. Here, we provide an overview of the history of RCTs and discuss the major challenges and limitations of current critical care RCTs, including overly optimistic effect sizes; unnuanced conclusions based on dichotomization of results; limited focus on patient-centred outcomes other than mortality; lack of flexibility and ability to adapt, increasing the risk of inconclusive results and limiting knowledge gains before trial completion; and inefficiency due to lack of re-use of trial infrastructure. We discuss recent developments in critical care RCTs and novel methods that may provide solutions to some of these challenges, including a research programme approach (consecutive, complementary studies of multiple types rather than individual, independent studies), and novel design and analysis methods. These include standardization of trial protocols; alternative outcome choices and use of core outcome sets; increased acceptance of uncertainty, probabilistic interpretations and use of Bayesian statistics; novel approaches to assessing heterogeneity of treatment effects; adaptation and platform trials; and increased integration between clinical trials and clinical practice. We outline the advantages and discuss the potential methodological and practical disadvantages with these approaches. With this review, we aim to inform clinicians and researchers about conventional and novel RCTs, including the rationale for choosing one or the other methodological approach based on a thorough discussion of pros and cons. Importantly, the most central feature remains the randomisation, which provides unparalleled restriction of confounding compared to non-randomised designs by reducing confounding to chance.

Keywords: Clinical trials, Critical care, Randomized clinical trials, Intensive care

Introduction

Randomised clinical trials (RCTs) fundamentally changed the practice of medicine, and randomisation is the gold standard for providing unbiased estimates of intervention effects [1]. Clinical trials have evolved, substantially, from the first described systematic comparison of dietary regimens 2500 years ago in Babylon, to the 1747 scurvy trial, the first double-blinded trial of patulin for the common cold conducted in the 1940's, and the establishment of modern ethical standards and regulatory frameworks following World War II (Fig. 1) [2, 3].

*Correspondence: andersgran@gmail.com
[1] Department of Intensive Care, Rigshospitalet, University of Copenhagen, Copenhagen University Hospital, Blegdamsvej 9, 2100 Copenhagen, Denmark
Full author information is available at the end of the article

🌱 Springer

While the fundamental concept of RCTs has remained relatively unchanged since then, the degree of collaboration has increased, and the largest trials have become larger [4]. Additionally, smaller RCTs assessing efficacy in narrow populations in highly controlled settings have been complemented with larger, more pragmatic RCTs in broader populations with less protocolisation of concomitant interventions, more closely resembling clinical practice [5]. Similarly, per-protocol-analyses assessing efficacy (i.e., effects of an intervention under ideal circumstances in patients with complete protocol adherence) have been complemented with intention-to-treat-analyses, assessing effectiveness under pragmatic circumstances in all randomised patients, regardless of protocol adherence (which may be affected by the intervention itself). This provides a better estimate of the actual effects of choosing one intervention over another in clinical practice [5]. A discussion and historical timeline of key critical care studies and RCTs is available elsewhere [6].

In this review, we outline the characteristics and common challenges of conventional RCTs in critical care, discuss potential improvements and novel design features followed by discussion of their potential limitations.

Common limitations and challenges of RCTs in critical care
RCTs are not without limitations, some related to the conventional design (i.e., a parallel, two-group, fixed-allocation-ratio RCT analysed with frequentist methods) and several to how many RCTs are designed and conducted. First, most critical care RCTs compare two interventions; while appropriate if only two interventions are truly of interest, oversimplifications may occur when two interventions, doses, or durations are chosen primarily to simplify trials. Second, sample size estimations for most RCTs enrolling critically ill patients in the intensive care unit (ICU) use overly optimistic effect sizes [7–10], leading to RCTs capable of providing firm evidence for very large effects, but unable to confirm or refute smaller, yet clinically relevant effects. Consequently, critical care RCTs are frequently inconclusive from a clinical perspective, and "absence of evidence interpreted as evidence of absence"-errors of interpretation [11] are common when RCTs are analysed using frequentist statistical methods and interpreted according to whether 'statistical significance' has been reached [8, 11]. Ultimately, this may lead to beneficial interventions being prematurely abandoned, and it has been argued that the conduct of clearly underpowered RCTs is unethical [12]. Third, critical care RCTs frequently focus on mortality [8]; while patient-important [13] and capable of capturing both desirable and undesirable effects, it conveys limited information [14], thus requiring large samples. Interventions may reduce
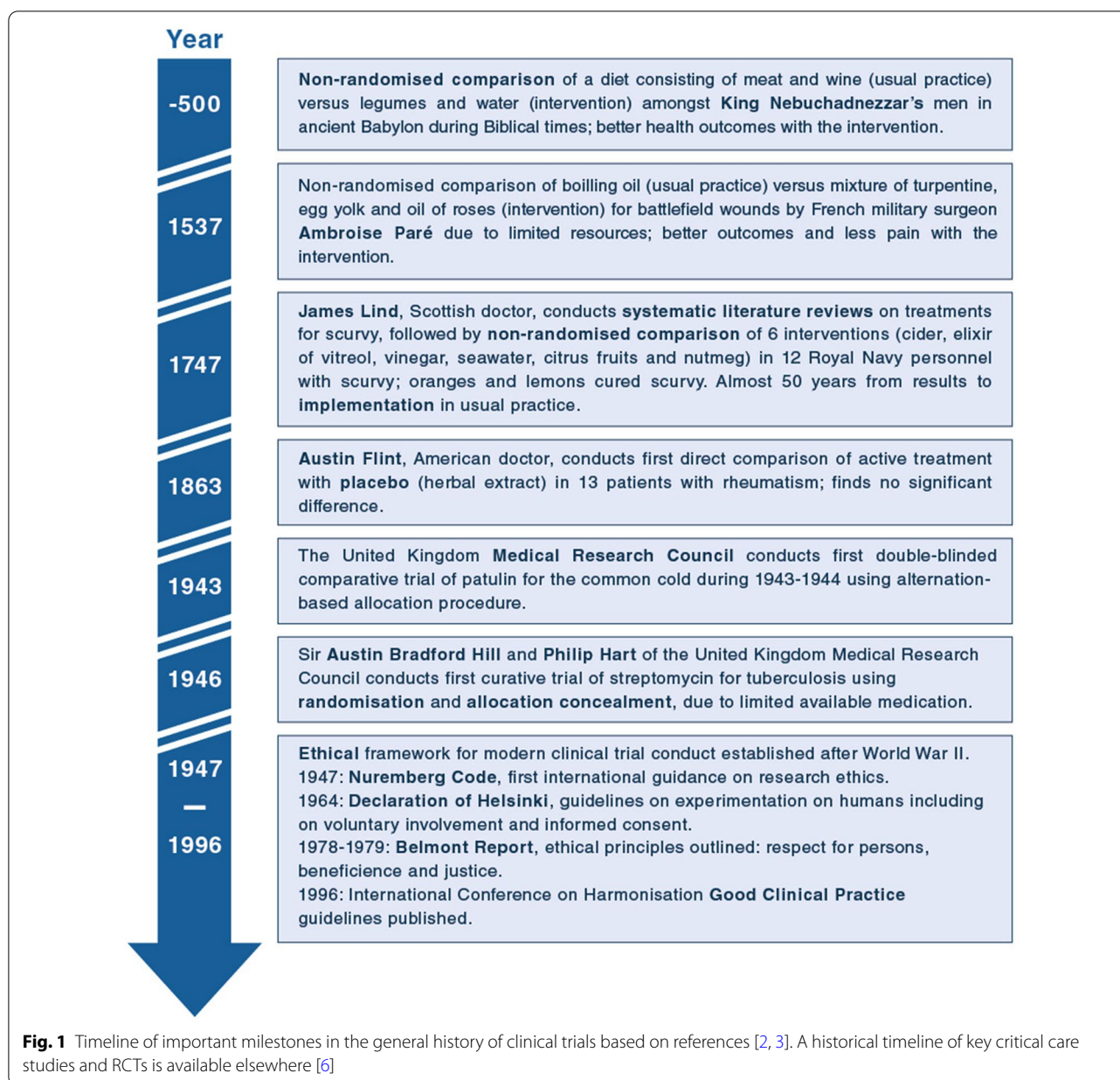
Take-home message

In this review, the primary challenges of conventional randomised clinical trials in critical care are discussed. This is followed by discussion of potential solutions and novel trial methods, including the challenges and potential disadvantages of using these methods.

morbidity or mortality due to one cause, but if assessed in patients at substantial risk of dying from other causes, differences may be difficult to detect [15]. Similarly, interventions may lead to negative intermediate outcomes and prolonged admission or increased treatment intensity, but not necessarily death [15]. Fourth, conventional RCT are inflexible. While one or few interim analyses may be conducted, they often rely on hard criteria for stopping [16]. Benefit or harm can, therefore, only be detected early if the effect is very large. Fifth, planning and initiating RCTs usually takes substantial time and funding, re-use of trial infrastructure is limited, data collection is mostly manual requiring substantial resources, and between-trial coordination is usually absent, increasing the risk of competing trials. Finally, even for conclusive RCTs, disseminating and implementing results into clinical practice requires substantial effort and time [17].

Larger trials, standardisation, meta-analyses and research programmes
The simple solution to inconclusive, underpowered RCTs is enrolling more patients, which requires more resources and international collaboration, while increasing external validity. Fewer, larger RCTs are more likely to produce conclusive evidence regarding important clinical questions than multiple, smaller trials, and are better to assess safety (including rare adverse events) if properly monitored. Thus, there may be a rationale for focussing on widely used interventions, such as the *Mega-ROX* RCT [18] that aims to compare oxygenation targets in 40,000 ICU patients to provide conclusive evidence for smaller effects than what previous RCTs have been able to confirm or reject [19–21].

An alternative to very large RCTs which for logistic, economic and administrative reasons are challenging is standardisation or harmonisation of RCT protocols, followed by pre-planned, prospective meta-analyses [22], which may also limit competition between trials. This was done for three large RCTs of early goal-directed therapy for septic shock, with results included in a conventional, trial-level meta-analysis with other RCTs and a prospectively planned individual patient-data meta-analysis [23, 24]. Other examples include prospective meta-analyses on systemic corticosteroids and interleukin-6-receptor antagonists for critically ill patients with

Fig. 1 Timeline of important milestones in the general history of clinical trials based on references [2, 3]. A historical timeline of key critical care studies and RCTs is available elsewhere [6]

coronavirus disease 2019 (COVID-19) [25, 26], and a prospective meta-trial including six RCTs of awake prone positioning for patients with COVID-19 and hypoxia, with separate logistics and infrastructure, but harmonised protocols and prospective analysis of combined individual participant-data [27].

Importantly, RCTs should ideally be conducted as part of complete research programmes (Fig. 2), with pre-clinical studies (e.g., in-vitro and animal studies), systematic reviews, and non-randomised studies and pilot/feasibility RCTs informing RCT designs, including selection of appropriate research questions, populations,

interventions and comparators, outcomes and realistic effect sizes. When RCTs are completed, results should be incorporated in updated systematic reviews and clinical practice guidelines to ease implementation [28], all considering relevant patient differences and effects of concomitant interventions. For example, the SUP-ICU programme included topical and systematic reviews summarising existing evidence, a survey describing preferences and indications for stress ulcer prophylaxis, and a cohort assessing prevalence, risk factors and outcomes of patients with gastrointestinal bleeding before the RCT was designed [29–33]. Following the RCT, results were
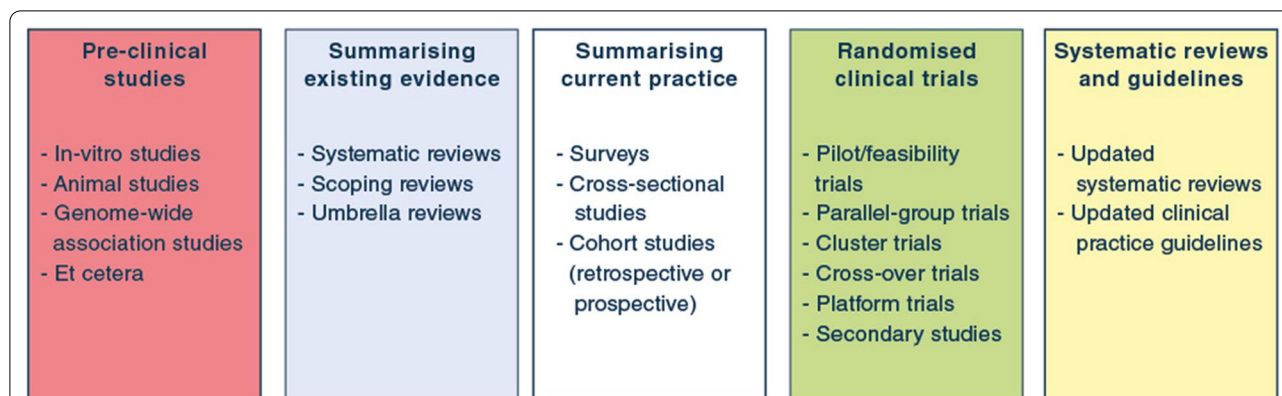
Fig. 2 Overview of different study types and their role in clinical research programmes. In general, pre-clinical studies can provide necessary background or laboratory knowledge that may be used to generate hypotheses later assessed in clinical trials. Summarising existing evidence prior to start of clinical studies is sensible, to identify knowledge gaps, avoid duplication of efforts, and inform further clinical studies. Surveys may identify existing beliefs, practices and attitudes towards further studies; cross-sectional studies and cohort studies can describe prevalence, outcomes, predictors/risk factors and current practice. Randomised clinical trials remain the gold standard for intervention comparisons but may also provide data for secondary studies not necessarily focussing on the randomised intervention comparison. Before randomised clinical trials aimed at assessing efficacy or effectiveness of an intervention are conducted, pilot/feasibility trials may be conducted to prepare larger trials and assess protocol delivery and feasibility. Following the conduct of a randomised clinical trial, relevant systematic reviews and clinical practice guidelines should be updated as necessary, to ease implementation of trial results into clinical practice. Of note, the process is not always linear and unidirectional, and different study types may be conducted at different temporal stages during a research programme. Translational research may incorporate pre-clinical and laboratory studies and clinical studies, including non-randomised cohort studies and randomised clinical trials. Similarly, clinical studies may be used to collect data or samples that are further analysed outside the clinical setting

incorporated in updated systematic reviews and clinical practice guidelines [33–35].
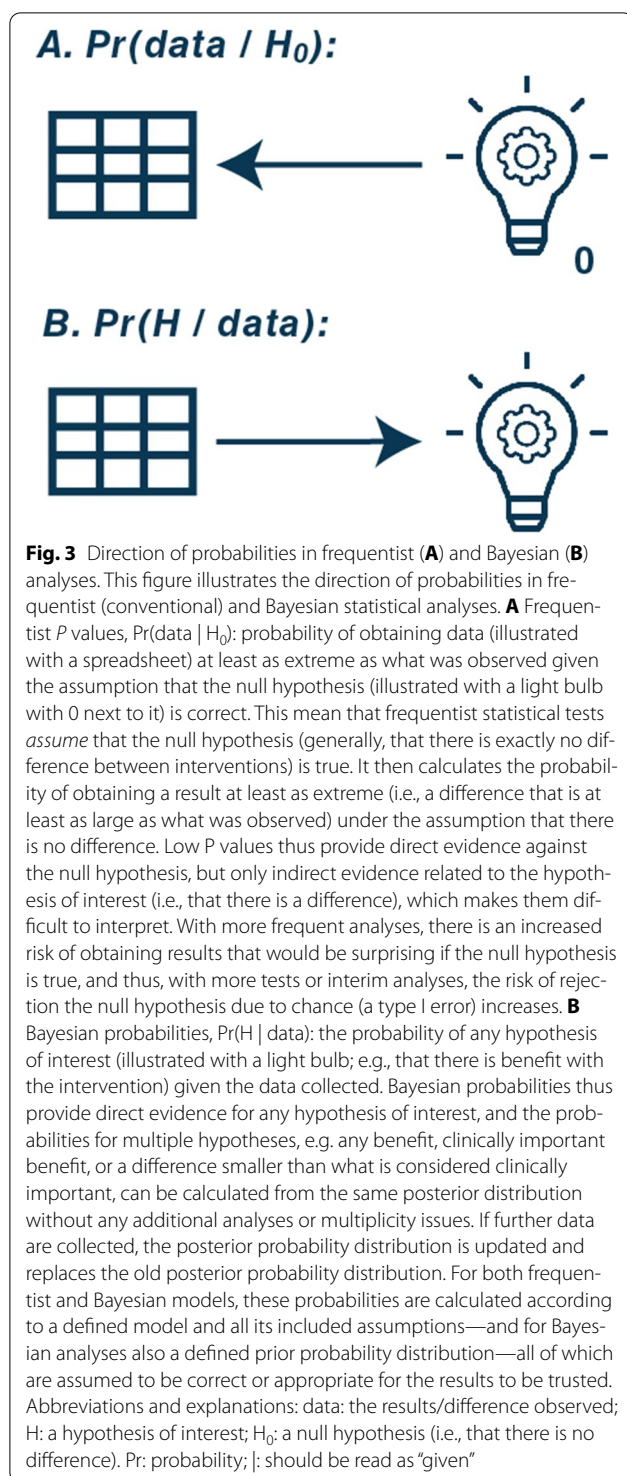
Outcome selection

Historically, most RCTs in critically ill patients have focussed on all-cause landmark mortality assessed at a single time-point [8]. As mortality in critically ill patients is high, it needs to be considered regardless of the outcome chosen. However, mortality conveys limited statistical information compared to more granular outcomes, as it only contains two possible values, i.e., death or alive regardless of health state [14, 36] and is thus insensitive to changes leading to other clinical improvements, e.g., quicker disease resolution or better functional outcomes in survivors. Thus, mortality requires large samples, and RCTs focussing on mortality are less frequently 'statistically significant' compared to RCTs focussing on other outcomes [8]. While mortality may be the most appropriate outcome in some trials, other outcomes should thus be considered [37]. During the COVID-19 pandemic, multiple RCTs focussed on more granular, higher-information outcomes such as days alive without life support or mechanical ventilation [38–40], which includes both mortality, resource use and illness durations. However, these outcomes are challenging due to different definitions, different handling of death, potentially opposing effects on mortality and the duration of life support in survivors, possibly greater risk of bias in unblinded trials,

and difficult statistical analysis [41–43]. Use of composite outcomes may increase power due to overall more events, but hamper interpretability as components of different importance to patients are weighted equally, and as interventions may affect individual components differently (e.g., increase intubation rates but decrease mortality) [44]. Finally, the development of core outcome sets may help in prioritisation and standardising outcome selection, allowing easier comparison and synthesis of RCT results [45].

Avoiding dichotomisation and embracing uncertainty

Most RCTs are planned and analysed using frequentist statistical methods, with results dichotomised as 'statistically significant' or not. Non-significant results are misinterpreted as evidence for no difference in approximately half of journal articles [46] and avoiding dichotomisations and abandoning the concept of statistical significance has been repeatedly discussed and recommended [46–49].

P values are calculated assuming that the null hypothesis is true (i.e., that there is exactly no difference, which is often implausible), and as they are indirect probabilities, they are hard to interpret (Fig. 3) [50]. As *P* values depend on both effect sizes and sample sizes, they will generally be small in large samples and large in small samples, regardless of the potential clinical importance of effects; thus, estimating effect sizes with uncertainty measures [i.e., confidence intervals (CIs)] may be preferable [51],

Fig. 3 Direction of probabilities in frequentist (**A**) and Bayesian (**B**) analyses. This figure illustrates the direction of probabilities in frequentist (conventional) and Bayesian statistical analyses. **A** Frequentist *P* values, Pr(data | H$_0$): probability of obtaining data (illustrated with a spreadsheet) at least as extreme as what was observed given the assumption that the null hypothesis (illustrated with a light bulb with 0 next to it) is correct. This mean that frequentist statistical tests *assume* that the null hypothesis (generally, that there is exactly no difference between interventions) is true. It then calculates the probability of obtaining a result at least as extreme (i.e., a difference that is at least as large as what was observed) under the assumption that there is no difference. Low P values thus provide direct evidence against the null hypothesis, but only indirect evidence related to the hypothesis of interest (i.e., that there is a difference), which makes them difficult to interpret. With more frequent analyses, there is an increased risk of obtaining results that would be surprising if the null hypothesis is true, and thus, with more tests or interim analyses, the risk of rejection the null hypothesis due to chance (a type I error) increases. **B** Bayesian probabilities, Pr(H | data): the probability of any hypothesis of interest (illustrated with a light bulb; e.g., that there is benefit with the intervention) given the data collected. Bayesian probabilities thus provide direct evidence for any hypothesis of interest, and the probabilities for multiple hypotheses, e.g. any benefit, clinically important benefit, or a difference smaller than what is considered clinically important, can be calculated from the same posterior distribution without any additional analyses or multiplicity issues. If further data are collected, the posterior probability distribution is updated and replaces the old posterior probability distribution. For both frequentist and Bayesian models, these probabilities are calculated according to a defined model and all its included assumptions—and for Bayesian analyses also a defined prior probability distribution—all of which are assumed to be correct or appropriate for the results to be trusted. Abbreviations and explanations: data: the results/difference observed; H: a hypothesis of interest; H$_0$: a null hypothesis (i.e., that there is no difference). Pr: probability; |: should be read as "given"

although CIs are frequently misinterpreted, too [50]. As misinterpretations are common [46, 50], increased education of clinicians and researchers is likely needed [48].

The issue with dichotomising results received attention following the publication of several important critical care RCTs with apparent discrepancies between statistical significance and clinical importance. The *EOLIA* RCT of extracorporeal membrane oxygenation (ECMO) in patients with severe acute respiratory distress syndrome (ARDS) concluded that "60-day mortality was not significantly lower with ECMO than with a strategy of conventional mechanical ventilation that included ECMO as rescue therapy." [52]. While technically correct, it may be considered overly reductionistic, as the conclusion was based on 60-day mortality rates of 35% (ECMO) vs. 46% (control) and a *P* value of 0.09 following a sample size calculation based on an absolute risk reduction of 20 percentage points [52]. Similarly, the *ANDROMEDA-SHOCK* RCT conducted in septic shock patients concluded that "a resuscitation strategy targeting normalization of capillary refill time, compared with a strategy targeting serum lactate levels, did not reduce all-cause 28-day mortality.", based on 28-day mortality rates of 34.9% vs. 43.4% and a *P* value of 0.06, following a sample size calculation based on a 15 percentage points absolute risk reduction [53]. Arguably, smaller effect sizes are clinically relevant in both cases.

There has been increased interest in supplementing or replacing conventional analyses with Bayesian statistical methods [37, 54, 55], which start with probability distributions expressing prior beliefs. Once data have been collected, these are updated to posterior probability distributions [56, 57]. Different prior distributions can be used, including uninformative-, vaguely informative-, evidence-based-, sceptic-, positive- or negative priors [58]. The choice of prior may be difficult and may potentially be abused to get the 'desired' results; typically, however, weakly informative, neutral priors, with minimal influence on the results are used for the primary Bayesian analyses of critical care RCTs, with sensitivity analyses assessing the influence of other priors [59–63]. If priors are transparently reported (and ideally pre-specified), assessing whether they are reasonable is fairly easy. Posterior probability distributions can be summarised in multiple ways. Credible intervals (CrIs) directly represent the most probable values (which is how frequentist CIs are often erroneously interpreted) [50, 57], and direct probabilities of any effect size can be calculated, i.e., the probability of any benefit (relative risk < 1.00), clinically important benefit (e.g., absolute risk difference > 2 percentage points) or practical equivalence (e.g., absolute risk difference between −2 and 2 percentage points) (Fig. 3).

In Bayesian re-analyses of EOLIA and ANDROMEDA-SHOCK, there were 96% and 98% probabilities of benefit with the interventions, respectively, using minimally informative or neutral priors [59, 60]; while thresholds for adopting interventions may vary

depending on resources/availability, preferences, and cost, these re-analyses led to more nuanced interpretations, with the use of multiple priors allowing readers to form their own context-dependent conclusions. Several Bayesian analyses have been conducted post hoc [59, 60, 64–66], sometimes motivated by apparently clinically important effect sizes that did not reach statistical significance, while others have been pre-specified [61, 62, 67], which is preferable as selection driven by trial results is thus avoided.

Nuanced interpretations avoiding dichotomisations are also possible using conventional, frequentist statistics [46–49]; however, assessments of statistical significance may be so ingrained in many clinicians, researchers and journal editors that more nuanced interpretation may be easier facilitated with alternative statistical approaches. Different evidence thresholds may be appropriate depending on the intervention, i.e., less certain evidence may be required when comparing commonly used and well-known interventions with similar costs and disadvantages, and more certain evidence may be required before implementing new, costly or burdensome interventions [68]. This is similar to how clinical practice guidelines consider the entire evidence base and the nature of the interventions being compared including costs, burden of implementation and patient preferences [1]. While claims of "no difference" based solely on lack of statistical significance should be avoided, clearly pre-defined thresholds may still be required for approving new interventions, for declaring trials "successful" and for limiting the risk of "spin" in conclusions. Thus, a nuanced set of standardised policy responses to more nuanced evidence

summaries may be warranted to ensure some standardisation of interpretation and implementation, while still considering differences in patient characteristics and preferences.

Average and heterogeneity of treatment effects

The primary RCT results generally represent the average treatment effects across all included patients, however, heterogeneity of treatment effects (HTE) [69, 70] in subpopulations are likely, and, despite being difficult to prove, have been suggested in multiple previous critical care RCTs [33, 64, 71–74]. A neutral average effect may represent benefit in some patients and harm in others (Fig. 4), and a beneficial average effect may differ in magnitude across subgroups, which could influence decisions to use the intervention [1, 75, 76]. It is sometimes assumed that the risk of adverse events is similar for patients at different risk of the primary outcome [70], which may affect the balance between benefits and harms of a treatment according to baseline risk, although this assumption may not always hold [77].

While large, pragmatic RCTs may be preferred for detecting clinically relevant average treatment effects, guiding overall clinical practice recommendations and for public healthcare, they have been criticised for including too heterogenous populations, often due to inclusion of general acutely ill ICU patients or ICU patients with broad syndromic conditions, i.e., sepsis or ARDS [78]. Even if present, HTE may be of limited importance if some patients benefit while others are mostly unaffected, if cost or burden of implementation is limited, or if some patients are harmed while others are mostly unaffected.
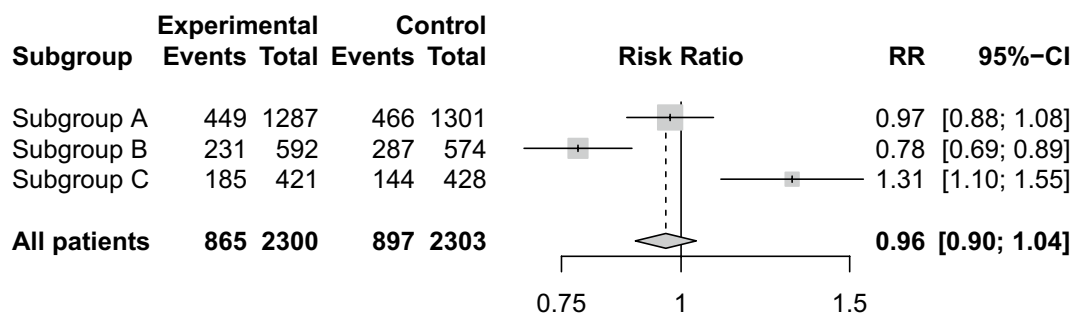

Fig. 4 Heterogeneity of treatment effects in clinical trial. Forest plot illustrating a fictive clinical trial enrolling 4603 patients. In this trial, the average treatment effect may be considered neutral with a relative risk (RR) of 0.96 and 95% confidence interval of 0.90–1.04 (or inconclusive, if this interval included clinically relevant effects). The trial population consists of three fictive subgroups with heterogeneity of treatment effects: A, with an intervention effect that is neutral (or inconclusive), similarly to the pooled result; B, with substantial benefit from the intervention; and C, with substantial harm from the intervention. If only the average intervention effect is assessed, it may be concluded – based on the apparent neutral overall result – that whether the intervention or control is used has little influence on patient outcomes, and it may be missed that the intervention provides substantial benefit in some patients and substantial harm in others. Similarly, an intervention with an overall beneficial effect may be more beneficial in some subgroups than others and may provide harm in some patients, and vice versa

Most RCT assess potential HTE by conducting conventional subgroup analyses despite important limitations [79]. As substantially more patients are required to assess subgroup differences than for primary analyses, most subgroup analyses are substantially underpowered and may miss clinically relevant differences [79]. In addition, larger numbers of subgroup analyses increase the risk of chance findings [79]. Conventional subgroup analyses assess one characteristic at a time, which may not reflect biology or clinical practice where multiple risk factors are often synergistic or additive [79], or where effect modifiers may be dynamic and change during illness course. Finally, conventional subgroup analyses frequently dichotomise continuous variables, which limits power [80] and makes assessment of gradual changes in responses difficult.

Alternative and better solutions for assessing HTE include predictive HTE analysis, where a prediction model incorporating multiple relevant clinical variables predictive of either the outcome or the change in outcomes with the intervention is used [77]; use of clustering algorithms and clinical knowledge to identify subgroups and distinct clinical pheno-/endotypes for syndromic conditions [64, 81, 82]; assessments of interactions with continuous variables without categorisation [63–65]; use of Bayesian hierarchical models, where subgroups effect estimates are partially pooled, limiting the risk of chance findings in smaller subgroups [63–65]; and adaptive enrichment [83, 84], discussed below. Improved and more granular analyses seem the most realistic way towards "personalised" medicine [77], but requires more data and thus overall larger RCTs. Regardless of the approach, appropriate caution should always be employed when interpreting subgroup and HTE analyses.

Adaptation

Adaptive trials are more flexible and can be more efficient than conventional RCTs [85], while being designed to have similar error rates. Adaptive trials often, but not always, use Bayesian statistical methods, which are well suited for continuous assessment of accumulating evidence [83, 86]. Adaptive trials can be adaptive in multiple ways [87]. First, pre-specified decision rules (for stopping for inferiority/superiority/equivalence/futility) allow trials to run without pre-specified sample sizes or to revise target sample sizes, thus allowing trials to run until just enough data have been accumulated. Expected sample sizes are estimated using simulation; if the expected baseline risks and effect sizes are incorrect, the final sample sizes will differ from expectations, but adaptive trials are still able to continue until sufficient evidence is obtained. Further, adaptive sample sizes are better suited for new diseases, where no or limited existing knowledge

complicates sample size calculations. For example, conducting the OSCAR RCT assessing high-frequency oscillation in ARDS using a Bayesian adaptive design could have reduced the number of patients and total deaths by > 15% [88]. Second, trials may be adaptive regarding the interventions assessed; multiple interventions or doses may be studied simultaneously or in succession, and the least promising may be dropped while assessment of better performing interventions continues until conclusive evidence has been obtained [83, 86]. This has been used for dose-finding trials, e.g., the *SEPSIS-ACT* RCT initially compared three selepressin doses to placebo, followed by selection of the best dose for further comparison [67], and the ongoing adaptive phase II/III Revolution trial [89], comparing antiviral drugs and placebo focussing on reducing viral loads in its first phases and increasing the number of days without respiratory support in the third phase. Similarly, interventions may be added during the trial, as in platform trials discussed below. Third, trials may use response-adaptive randomisation to update allocation ratios based on accumulating evidence, thereby increasing the chance that patients will be allocated to more promising interventions, despite not having reached conclusiveness yet. This can increase efficiency in some situations, but also decrease it, as in two-armed RCT and some multi-armed RCTs [90, 91]. Thus, it has been argued that while response-adaptive randomisation may benefit *internal* patients, it may not always be preferable, as it can lead to slower discovery of interventions that can benefit patients external to the trial in some cases [91, 92]. Finally, trials may use adaptive enrichment to adapt/restrict inclusion criteria to focus on patients more likely to benefit, or use different allocation ratios for different subpopulations [84, 87].

Platform trials

Platform trials are RCTs that instead of focussing on single intervention comparisons focus on a disease or condition and assess multiple interventions according to a master protocol [83, 93]. Platform trials may run perpetually, with interventions added or dropped continuously [83, 94] and often employ multiple adaptive features and probabilistic decision rules [83, 93]. Interventions assessed can be nested in multiple domains, e.g., *REMAP-CAP* assesses interventions in patients with severe community-acquired pneumonia in several domains including antibiotics, corticosteroids, and immune-modulating therapies. By assessing multiple interventions simultaneously and by re-using controls for comparisons with multiple interventions, platform trials can be more efficient than sequential two-armed comparisons and can be more efficient than simpler adaptive trials [94, 95].

Adaptive platform trials are capable of "learning while doing", and potentially allow tighter integration of clinical research and clinical practice, i.e., a better exploration–exploitation trade-off (learning versus doing based on existing knowledge) [83, 96]. If response-adaptive randomisation is used, probabilities of allocation to potentially superior interventions increases as evidence is accumulated, and interventions that are deemed superior may immediately become implemented as standard of care by becoming the new control group [83]. Thus, implementation of results into practice – at least in participating centres—may become substantially faster. While platform trials have only recently been used in critically ill patients, the *RECOVERY* and *REMAP-CAP* trials have led to substantial improvements in the treatment of patients with COVID-19 within a short time-frame [38, 97–99], although this may not only be explained by the platform design, but also the case load and urgency of the situation.

Comparable to how data from multiple conventional RCTs may be prospectively planned to be analysed together, data from multiple platform trials may be combined in multiplatform trials with similar benefits and challenges as individual platform trials and standardisation across individual, conventional RCTs [100].

Further embedding of RCTs into clinical practice

In addition to the possible tighter integration between research and clinical practice that may come with adaptive platform trials and ultimately may lead to learning healthcare systems [83], integration may be increased in other ways. Trials may be embedded in electronic health records, where automatic integration may lead to substantial logistic improvements regarding data collection, integration of randomisation modules, and alerts about potentially eligible patients. This may improve logistics and data collection and facilitate closer integration between research and practice [61, 83]. Similarly, RCTs may use data already collected in registers or clinical databases, substantially decreasing the data-collection burden, as has been done in, e.g., the *PEPTIC* cluster-randomised register-embedded trial [74]. Finally, fostering an environment where clinical practice and clinical research are tightly integrated and where enrolment in clinical trials is considered an integral part of clinical practice in individual centres by clinicians, patients and relatives may lead to faster improvements of care for all patients.

Limitations and challenges

While the methods discussed may mitigate some challenges of conventional RCTs, they are not without limitations (Table 1). First, larger trials come with challenges regarding logistics, regulatory requirements (including approvals, consent procedures, and requirements for reporting adverse events), economy, collaboration, between-centre heterogeneity in other interventions administered, and potential challenges related to academic merits. Second, standardisation and meta-analyses may require compromises or increased data-collection burden in some centres or may not be possible due to between-trial differences. Third, while complete research programmes may lead to better RCTs, they may not be possible in, e.g., emergency situations such as pandemics caused by new diseases. Fourth, using outcomes other than mortality comes with difficulties relating to statistical analysis, how death is handled and possibly interpretation, and mortality should not be abandoned for outcomes that are not important to patients. Fifth, while avoiding dichotomisation of results and using Bayesian methods has some advantages, it may lead to larger differences in how evidence is interpreted and possibly lower thresholds for accepting new evidence if adequate caution is not employed. In addition, switching to Bayesian methods requires additional education of clinicians, researchers and statisticians, and specification of priors and estimating required sample sizes adds complexity. Sixth, while improved analyses of HTE have benefits compared to conventional subgroup analyses, the risk of chance findings and lack of power remains. Finally, adaptive and platform trials come with logistic and practical challenges as listed in Table 1 and discussed below.

As adaptive and platform trials are substantially less common than more conventional RCTs, there is less methodological guidance and interpretation may be more difficult for readers. Fortunately, several successful platform trials have received substantial coverage in the critical care community [64, 99], and an extension for the Consolidated Standards of Reporting Trials (CONSORT) statement for adaptive trials was recently published [101]. Planning adaptive and platform trials comes with additional logistic and financial challenges related to the current project-based funding model, which is better suited for fixed-size RCTs [83, 85, 93]. While adaptive trials are more flexible, large samples may still be required to firmly assess all clinically relevant effect sizes, which may not always be feasible. In addition, statistical simulation is required instead of simple sample size estimations [83, 94]. Further, the regulatory framework for adaptive and platform trials is less well-developed than for conventional RCTs, and regulatory approvals may thus be more complex and time-consuming [83].

There are also challenges with the adaptive features, and careful planning is necessary to avoid aggressive adaptations to random, early fluctuations. Initial "burn-in" phases where interventions are not compared until a

Table 1 Methodological features that may improve clinical trials benefits and challenges

Feature	Benefits and advantages	Challenges and disadvantages
Larger trials (increased sample sizes)	Decreased uncertainty, increased precision Easier to detect potential subgroup differences Less chance of inconclusive results (i.e., greater precision and less uncertainty); results from fewer large RCTs are easier to compare than results from many smaller RCTs Easier to address safety concerns if properly monitored, as larger trials have higher chances of detecting rare adverse events Increased generalisability/external validity in multicentre/international RCTs	Economic: trial cost and optimal use of overall research resources Collaboration: increased workload coordinating, different regulatory requirements including different handling of consent procedures and reporting of adverse events, challenges with coordination due to language and time zone differences Comparability: potential differences in standard of care/available resources in international trials Academic challenges: less individually led projects due to increased collaboration – group authorships may be less attractive in settings where individual author positions are valued (e.g., grant applications)
Standardisation and meta-analyses	Increased comparability/less heterogeneity Less competition between trials Meta-analysis may be more sensible – less statistical inconsistency may lead to more precise results Prospective meta-analyses or meta-trials may provide quicker answers than individual conventional RCTs and meta-analyses, especially if trialists share data earlier, and adequate certainty is obtained before individual trials finish	Agreement between investigators on the design and variables could be challenging and time-consuming; compromises may be necessary for standardisations; core outcome sets may improve this Data not routinely collected in one setting may be required due to standardisation, potentially increasing workload in some centres If adequate standardisation is not possible, comparisons in meta-analyses may be difficult Differences in populations, interventions, comparators, outcomes, concurrent treatments and changes over time may hamper interpretation of meta-analyses
Research programmes	Complete research programmes including multiple study types may lead to better RCTs focussing on more relevant questions Evidence synthesis prior to trial conduct puts trials into context and may help identify the largest knowledge gaps or where new trials are not necessary	Research programmes may require substantial resources and time until an eventual trial can start; in most situations this will be sensible, but may not be possible during pandemics or emergencies and may require additional resources and funding
Outcome choices	Choosing non-dichotomous or non-mortality outcomes carrying more information may lead to more efficient or conclusive trials and smaller sample size requirements Outcomes with more levels than just dead/alive may convey important information on how well survivors fare	Definition and handling of death is challenging, including appropriate "weighting" of death, and clinical interpretation if mortality is treated in a special manner (e.g., if days alive without life support is analysed as an ordinal variable with death treated as worse than 0 days) Many non-mortality patient-important outcomes have skewed distributions complicating many common statistical (parametric) analyses and estimations of differences on an interpretable scale (including in meta-analyses) Difficulties in interpretation if effects on mortality and other parts of the outcome are in different directions (for composite outcomes and days alive without life support and similar outcomes) Risk of choosing less patient-important outcomes or surrogate outcomes

Table 1 (continued)

Feature	Benefits and advantages	Challenges and disadvantages
Avoiding dichotomisation of results, probabilistic interpretations	Nuanced conclusions; assessing evidence as a continuum avoids risk of incorrect "absence of evidence interpreted as evidence of absence" errors Using Bayesian methods allow incorporation of previous results or scepticism and easier propagation of uncertainty to subsequent calculations The same level of evidence may not be required to change clinical practice for all interventions—this depends on price, risk of adverse events, availability, character of intervention, invasiveness, etc.; these considerations apply to both trials and clinical practice guidelines	Probabilistic interpretations do not solve the primary issues of many trials; lack of dichotomisation does not in itself increase the certainty of evidence While conventional significance thresholds are arbitrary, they are widely used; changing methods may lead some researchers to opt for less strict thresholds or allow increased "spin" in conclusions. Disagreements in interpretation may increase if there is no standard threshold and pre-specified criteria for success for e.g. approving new interventions and standardised policy responses may be warranted Non-dichotomous and more detailed interpretations of trial results may be more difficult to communicate to non-researchers and non-experts Prior selection in Bayesian analyses adds additional complexity; results may be unduly influenced by strong(er) priors not shared by other researchers. Sensible priors (often non- or weakly informative priors are used in the primary Bayesian analyses of critical care trials), transparently reported, ideally pre-specified, and with adequate sensitivity analyses performed is warranted
Improved HTE analyses	Predictive HTE analyses and other approaches considering multiple patient characteristics simultaneously or overall risk may better reflect clinical reality than one-variable-at-a-time subgroup analyses Hierarchical models may limit the risk of exaggerated results and chance findings in smaller subgroups and increase precision due to borrowing of information Assessment of HTE according to variables of interest on the continuous scale may better detect dose–response relationships than categorised subgroup analyses	Subgroup or HTE analyses, regardless of approach, generally requires more patients—trials may still be underpowered to detect differences The more analyses conducted, the greater risk of chance findings—this may be mitigated, but not completely solved, by the discussed approaches Requires careful consideration of whether HTE analyses should be conducted on the absolute or relative scales; when the baseline risk differs between groups, there will always be HTE on either the relative or the absolute scale (often, intervention effects are most consistent on the relative scale)

Table 1 (continued)

Feature	Benefits and advantages	Challenges and disadvantages
Adaptation	Adaptive sample sizes/stopping rules may lead to optimally sized trials, more likely to reach conclusive evidence Adaptive arm adding/dropping may increase overall trial efficiency Adaptive randomisation may increase chance of getting better interventions in some situations, which may make trial participation more attractive to patients Adaptive enrichment may enable trials to better detect differences in responses and tailor interventions to different subpopulations or phenotypes	Logistic and economic challenges in planning and funding trials without fixed sample sizes; alternative financing models may be necessary Planning may be more difficult; instead of simple sample size calculations, advanced statistical simulation may be necessary to estimate required sizes and risk of random errors, requiring increased collaboration with statisticians and increased training of clinician-researchers Pre-specified criteria for stopping/adaptation necessary; may be difficult to define Adaptation requires more real-time data collection and verification, increasing data registration burden on individual sites Adaptations may be complex to implement and communicate Outcomes with longer follow-up lead to slower adaptation compared to shorter-term outcomes, which may add additional complexity. Consequently, the use of shorter-term outcomes to guide adaptive trials instead of the outcome of primary interest may be considered in some situations Risk of adaptations based on chance findings/fluctuations may require restraints of adaptation to avoid random errors, which is difficult to plan and handle While adaptive trials may be more likely to reach conclusive evidence if continued until a stopping rule is reached, they may need to be substantially larger to confirm or refute all clinically relevant effects (as is the case for conventional trials, too)
Adaptive platform trials	Increased efficiency, and potentially similar advantages as for adaptive conventional trials May decrease time to clinical adaptation and enable "learning while doing" Reuse of trial infrastructure and embedding in electronic health records and clinical practice may increase efficiency and decrease cost Potential improvement of informed consent procedures compared to consent when co-enrolment in multiple trials occurs Familiarity and consistency with a common platform design may be easier in practice than repeated conduction of independent RCTs	Same challenges as adaptive trials in general Potential regulatory issues; less well-known design may complicate approvals May take longer time to setup and implement than regular trials More complex – may be more difficult to implement and train staff, more difficult to explain to patients/potential complication of consent procedures, relatives and other stakeholders, may be more difficult to work with for non-researcher clinicians Standards for conducting and reporting less developed; may be more difficult to report and explain results Additional complexity with time drift/temporal variation and response-adaptive randomisation and potential re-use of non-concurrent controls requires adequate statistical handling to avoid bias Potential challenges with workload/stress of perpetual trials

Table 1 (continued)

Feature	Benefits and advantages	Challenges and disadvantages
Embedding trials in clinical practice and registers	Tighter integration of clinical practice and clinical trials may lead to faster improvements in patient care Embedding clinical trials in electronic health records may reduce data-collection burden and cost and alert clinicians and researchers of eligible patients and clinical events Register-based trials (including register-based cluster-randomised trials) may reduce data-collection burden and trial cost by using clinical registers already in place	Register-based data-collection may not be as easily standardised without changing individual registers; compromises based on availability in registers may be necessary Embedding trials in registers or electronic health records poses additional challenges with different electronic health record software and across borders Data quality and completeness in registers may not be as good as when data are prospectively collected for all variables Limited long-term outcome data generally available in registers due to additional complexity of data collection

HTE heterogeneity of treatment effects; *RCT* randomised clinical trial

sufficient number of enrolled patients can be used, as can more restrictive rules for response-adaptive randomisation and arm dropping early in the trial [94]. Simulation may be required to ensure that the risk of stopping due to chance is kept at an acceptable level, analogous to alpha-spending functions in conventional, frequentist trials [102]. Temporal changes in case-mix or concomitant interventions used may influence results in all RCTs, but is complicated further if adaptive randomisation or arm dropping/adding is used, thus requiring additional consideration, especially if patients randomised at earlier stages are re-used for comparisons with more recently introduced interventions [83]. Finally, comparisons with non-concurrent controls may affect interpretation and introduce bias if inappropriately handled [103].

Adaptations require continues protocol amendments and additional resources to implement and communicate, and may require additional training when new interventions are added [104]. Finally, while platform trials come with potential logistic and efficiency benefits, they may be more time-consuming initially and lack of a clear-cut "finish-line" may stress involved personnel [105], although familiarity and consistency may also have the opposite effect once implemented compared with repeated initiation, running and closure of consecutive, independent RCTs.

Future directions

We expect that the discussed methodological features will become more common in future critical care RCTs, and that this will improve efficiency and flexibility, and may help answer more complex questions. These methods come with challenges, though, and conventional RCTs may be preferred for simple, straightforward comparisons. Some challenges may be mitigated as these designs become more familiar to clinicians and researchers, and as additional methodological guidance is developed. We expect the future critical care RCT landscape to be a mix of relatively conventional RCTs and more advanced, adaptive trials. We propose that researchers consider the optimal methodological approach carefully when planning new RCTs. While different designs may be preferable in different situations, the choice should be based on careful thought instead of convenience or tradition, and more advanced approaches may be necessary in some situations to move critical care RCTs and practice forward.

Conclusion

In this review, we have discussed challenges and limitations of conventional RCTs, along with recent developments, novel methodological approaches and their advantages and potential disadvantages. We expect

critical care RCTs to evolve and improve in the coming years. At its core, however, the most central feature of any RCT remains the randomisation itself, which provides unparalleled protection against confounding. Consequently, the RCT remains the gold standard for comparing different interventions in critical care and beyond.

Author details

[1] Department of Intensive Care, Rigshospitalet, University of Copenhagen, Copenhagen University Hospital, Blegdamsvej 9, 2100 Copenhagen, Denmark. [2] Department of Medicine, McMaster University, Hamilton, Canada. [3] Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Canada. [4] Department of Intensive Care Medicine, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands. [5] Julius Center for Health Sciences and Primary Care, Utrecht, The Netherlands. [6] UPMC and University of Pittsburgh Schools of the Health Sciences, Pittsburgh, PA, USA. [7] Intensive Care Unit, Federal University of Sao Paulo, Sao Paulo, Brazil. [8] Critical Care Division, The George Institute for Global Health and UNSW, Sydney, Australia. [9] Malcolm Fisher Department of Intensive Care, Royal North Shore Hospital, Sydney, Australia. [10] Regional Intensive Care Unit, Royal Victoria Hospital, Belfast, Northern Ireland, UK. [11] Department of Anaesthesiology, Critical Care and Pain, Tata Memorial Hospital, Homi Bhabha National Institute, Mumbai, India. [12] Critical Care Department, Saint-Louis Hospital, Paris, France. [13] Intensive Care National Audit and Research Centre, London, UK. [14] Medical Research Institute of New Zealand, Wellington, New Zealand. [15] Intensive Care Unit, Wellington Regional Hospital, Wellington, New Zealand. [16] Australian and New Zealand Intensive Care Research Centre, Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia. [17] Department of Critical Care, University of Melbourne, Parkville, VIC, Australia.

Author contributions

Conceptualisation: AG and MHM. Writing—original draft: AG and MHM. Writing – review and editing: all authors.

Declarations

Conflicts of interest

The Department of Intensive Care at Copenhagen University Hospital—Rigshospitalet (AG, AP, MHM) has received grants from the Novo Nordisk Foundation, Pfizer, Fresenius Kabi and Sygeforsikringen "danmark" outside the submitted work. The University Medical Center Utrecht (LD) has received grants from the European Commission (Rapid European COVID-19 Emergency research Response (RECOVER) Grant number H2020 – 101003589; European Clinical Research Alliance on Infectious Diseases (ECRAID) Grant number H2020-965313) and the Dutch funder ZonMW (ANAkinra voor de behandeling van CORonavirus infectious disease 2019 op de Intensive Care (ANACOR-IC) Grant Number 10150062010003) for REMAP-CAP. FGZ has received grants for investigator initiated clinical trials from Ionis Pharmaceuticals (USA) and Bactiguard (Sweden), all unrelated to this work. The Critical Care Division, The George Institute for Global Health (NEH) has received grants from Baxter, CSL, and Fresenius Kabi outside the submitted work. EA declares having received fees for lectures from Alexion, Sanofi, Baxter, and Pfizer. His institution has received research grants from Fisher&Payckle, MSD and Baxter.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Granholm A, Alhazzani W, Møller MH (2019) Use of the GRADE approach in systematic reviews and guidelines. Br J Anaesth 123:554–559
2. Nellhaus EM, Davies TH (2017) Evolution of clinical trials throughout history. Marshall J Med. https://doi.org/10.18590/mjm.2017.vol3.iss1.9
3. Baron J (2012) Evolution of clinical research: a history before and beyond James Lind. Perspect Clin Res 1:6–10
4. Anthon CT, Granholm A, Perner A, Laake JH, Møller MH (2019) Overall bias and sample sizes were unchanged in ICU trials over time: a meta-epidemiological study. J Clin Epidemiol 113:189–199
5. Ford I, Norrie J (2016) Pragmatic trials. N Engl J Med 375:454–463
6. Finfer S, Cook D, Machado FR, Perner A (2021) Clinical research: from case reports to international multicenter clinical trials. Crit Care Med 49:1866–1882
7. Ridgeon EE, Bellomo R, Aberegg SK et al (2017) Effect sizes in ongoing randomized controlled critical care trials. Crit Care 21:132
8. Harhay MO, Wagner J, Ratcliffe SJ et al (2014) Outcomes and statistical power in adult critical care randomized trials. Am J Respir Crit Care Med 189:1469–1478
9. Cuthbertson BH, Scales DC (2020) "Paying the Piper": the downstream implications of manipulating sample size assumptions for critical care randomized control trials. Crit Care Med 48:1885–1886
10. Abrams D, Montesi SB, Moore SKL et al (2020) Powering bias and clinically important treatment effects in randomized trials of critical illness. Crit Care Med 48:1710–1719
11. Altman DG, Bland JM (1995) Statistics notes: absence of evidence is not evidence of absence. BMJ 311:485
12. Altman DG (1980) Statistics and ethics in medical research III: how large a sample? Br Med J 281:1336–1338
13. Gaudry S, Messika J, Ricard JD et al (2017) Patient-important outcomes in randomized controlled trials in critically ill patients: a systematic review. Ann Intensive Care 7:28
14. Roozenbeek B, Lingsma HF, Perel P et al (2011) The added value of ordinal analysis in clinical trials: An example in traumatic brain injury. Crit Care 15:R127
15. Veldhoen RA, Howes D, Maslove DM (2020) Is Mortality a useful primary end point for critical care trials? Chest 158:206–211
16. Stallard N, Todd S, Ryan EG, Gates S (2020) Comparison of Bayesian and frequentist group-sequential clinical trial designs. BMC Med Res Methodol 20:4
17. Grol R, Grimshaw J (2003) From best evidence to best practice: effective implementation of change in patients' care. Lancet 362:1225–1230
18. Australian and New Zealand Intensive Care Society (2021) MEGA-ROX (ANZICS CTG endorsed study). https://www.anzics.com.au/current-active-endorsed-research/mega-rox/. Accessed 05 Apr 2021
19. The ICU ROX Investigators and the Australian and New Zealand Intensive Care Society Clinical Trial Group (2020) Conservative oxygen therapy during mechanical ventilation in the ICU. N Engl J Med 382:989–998
20. Barbateskovic M, Schjørring O, Krauss SR et al (2019) Higher versus lower fraction of inspired oxygen or targets of arterial oxygenation for adults admitted to the intensive care unit. Cochrane Database Syst Rev. https://doi.org/10.1002/14651858.CD012631.pub2
21. Schjørring OL, Klitgaard TL, Perner A et al (2021) Lower or higher oxygenation targets for acute hypoxemic respiratory failure. N Engl J Med 384:1301–1311
22. Reade MC, Delaney A, Bailey MJ et al (2010) Prospective meta-analysis using individual patient data in intensive care medicine. Intensive Care Med 36:11–21
23. Angus DC, Barnato AE, Bell D et al (2015) A systematic review and meta-analysis of early goal-directed therapy for septic shock: the ARISE, process and promise investigators. Intensive Care Med 41:1549–1560
24. The PRISM Investigators (2017) Early, goal-directed therapy for septic shock: a patient-level meta-analysis. N Engl J Med 376:2223–2234

25. The WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group (2020) Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: a meta-analysis. JAMA 324:1330–1341

26. The WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group (2021) Association between administration of IL-6 antagonists and mortality among patients hospitalized for COVID-19: a meta-analysis. JAMA 325:499–518

27. Ehrmann S, Li J, Ibarra-Estrada M et al (2021) Awake prone positioning for COVID-19 acute hypoxaemic respiratory failure: a randomised, controlled, multinational, open-label meta-trial. Lancet Respir Med. https://doi.org/10.1016/s2213-2600(21)00356-8

28. Alhazzani W, Møller MH, Cote EB, Citerio G (2019) Intensive care medicine rapid practice guidelines (ICM-RPG): paving the road of the future. Intensive Care Med 45:1639–1641

29. Krag M, Perner A, Wetterslev J, Møller MH (2013) Stress ulcer prophylaxis in the intensive care unit: Is it indicated? A topical systematic review. Acta Anaesthesiol Scand 57:835–847

30. Krag M, Perner A, Wetterslev J, Wise MP, Møller MH (2014) Stress ulcer prophylaxis versus placebo or no prophylaxis in critically ill patients: a systematic review of randomised clinical trials with meta-analysis and trial sequential analysis. Intensive Care Med 40:11–22

31. Krag M, Perner A, Wetterslev J et al (2015) Stress ulcer prophylaxis in the intensive care unit: an international survey of 97 units in 11 countries. Acta Anaesthesiol Scand 259:576–585

32. Krag M, Perner A, Wetterslev J et al (2015) Prevalence and outcome of gastrointestinal bleeding and use of acid suppressants in acutely ill adult intensive care patients. Intensive Care Med 41:833–845

33. Krag M, Marker S, Perner A et al (2018) Pantoprazole in patients at risk for gastrointestinal bleeding in the ICU. N Engl J Med 379:2199–2208

34. Barbateskovic M, Marker S, Granholm A et al (2019) Stress ulcer prophylaxis with proton pump inhibitors or histamin-2 receptor antagonists in adult intensive care patients: a systematic review with meta-analysis and trial sequential analysis. Intensive Care Med 45:143–158

35. Ye Z, Blaser AR, Lytvyn L et al (2020) Gastrointestinal bleeding prophylaxis for critically ill patients: a clinical practice guideline. BMJ 368:l6722

36. Harrell F (2020) Statistical thinking. information gain from using ordinal instead of binary outcomes. https://www.fharrell.com/post/ordinal-info/. Accessed 20 Apr 2021

37. Harhay MO, Casey JD, Clement M et al (2020) Contemporary strategies to improve clinical trial design for critical care research: insights from the First Critical Care Clinical Trialists Workshop. Intensive Care Med 46:930–942

38. The Writing Committee for the REMAP-CAP Investigators (2020) Effect of hydrocortisone on mortality and organ support in patients with severe COVID-19: the REMAP-CAP COVID-19 corticosteroid domain randomized clinical trial. JAMA 324:1317–1329

39. Munch MW, Meyhoff TS, Helleberg M et al (2021) Low-dose hydrocortisone in patients with COVID-19 and severe hypoxia: the COVID STEROID randomised, placebo-controlled trial. Acta Anaesthesiol Scand. https://doi.org/10.1111/aas.13941

40. The COVID STEROID 2 Trial Group (2021) 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. https://doi.org/10.1001/jama.2021.18295

41. Granholm A, Kaas-Hansen BS, Kjær MN et al (2021) Patient-important outcomes other than mortality in recent ICU trials: protocol for a scoping review. Acta Anaesthesiol Scand 65:1002–1007

42. Bodet-Contentin L, Frasca D, Tavernier E, Feuillet F, Foucher Y, Giraudeau B (2018) Ventilator-free day outcomes can be misleading. Crit Care Med 46:425–429

43. Yehya N, Harhay MO, Curley MAQ, Schoenfeld DA, Reeder RW (2019) Reappraisal of ventilator-free days in critical care research. Am J Respir Crit Care Med 200:828–836

44. Irony TZ (2017) The "Utility" in composite outcome measures: measuring what is important to patients. JAMA 318:1820–1821

45. Blackwood B, Marshall J, Rose L (2015) Progress on core outcome sets for critical care research. Curr Opin Crit Care 21:439–444

46. Amrhein V, Greenland S, McShane B (2019) Scientists rise up against statistical significance. Nature 567:305–307

47. Sterne JA, Smith GD (2001) Sifting the evidence: what's wrong with significance tests? BMJ 322:226–231

48. Li G, Walter SD, Thabane L (2021) Shifting the focus away from binary thinking of statistical significance and towards education for key stakeholders: revisiting the debate on whether it's time to de-emphasize or get rid of statistical significance. J Clin Epidemiol 137:104–112

49. Wasserstein RL, Lazar NA (2016) The ASA's statement on p-values: context, process, and purpose. Am Stat 70:129–133

50. Greenland S, Senn SJ, Rothman KJ et al (2016) Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 31:337–350

51. Dunkler D, Haller M, Oberbauer R, Heinze G (2020) To test or to estimate? P-values versus effect sizes. Transpl Int 33:50–55

52. Combes A, Hajage D, Capellier G et al (2018) Extracorporeal membrane oxygenation for severe acute respiratory distress syndrome. N Engl J Med 378:1965–1975

53. Hernández G, Ospina-Tascón GA, Damiani LP et al (2019) Effect of a resuscitation strategy targeting peripheral perfusion status vs serum lactate levels on 28-day mortality among patients with septic shock. JAMA 321:654–664

54. Zampieri FG, Casey JD, Shankar-Hari M, Harrell FE, Harhay MO (2021) Using Bayesian methods to augment the interpretation of critical care trials. An overview of theory and example reanalysis of the alveolar recruitment for acute respiratory distress syndrome trial. Am J Respir Crit Care Med 203:543–552

55. Goligher EC, Zampieri F, Calfee CS, Seymour CW (2020) A manifesto for the future of ICU trials. Crit Care 24:686

56. Bendtsen M (2018) A gentle introduction to the comparison between null hypothesis testing and Bayesian analysis: Reanalysis of two randomized controlled trials. J Med Internet Res 20:e10873

57. Kruschke JK (2015) Doing bayesian data analysis, 2nd edn. Academic Press, London

58. Sung L, Hayden J, Greenberg ML, Koren G, Feldman BM, Tomlinson GA (2005) Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. J Clin Epidemiol 58:261–268

59. Goligher EC, Tomlinson G, Hajage D et al (2018) Extracorporeal membrane oxygenation for severe acute respiratory distress syndrome and posterior probability of mortality benefit in a Post Hoc Bayesian analysis of a randomized clinical trial. JAMA 320:2251–2259

60. Zampieri FG, Damiani LP, Bakker J et al (2020) Effect of a resuscitation strategy targeting peripheral perfusion status vs serum lactate levels on 28-day mortality among patients with septic shock: a Bayesian reanalysis of the ANDROMEDA-SHOCK trial. Am J Respir Crit Care Med 201:423–429

61. Angus DC, Berry S, Lewis RJ et al (2020) The REMAP-CAP (randomized embedded multifactorial adaptive platform for community-acquired pneumonia) study. Rationale and design. Ann Am Thorac Soc 17:879–891

62. Granholm A, Munch MW, Myatra SN et al (2021) 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. https://doi.org/10.1007/s00134-021-06573-1

63. Klitgaard TL, Schjørring OL, Lange T et al (2021) Lower versus higher oxygenation targets in ICU patients with severe hypoxaemia: secondary Bayesian analyses of mortality and heterogeneous treatment effects in the HOT-ICU trial. Br J Anaesth. https://doi.org/10.1016/j.bja.2021.09.010

64. Zampieri FG, Costa EL, Iwashyna TJ et al (2019) Heterogeneous effects of alveolar recruitment in acute respiratory distress syndrome: a machine learning reanalysis of the Alveolar Recruitment for Acute Respiratory Distress Syndrome Trial. Br J Anaesth 123:88–95

65. Granholm A, Marker S, Krag M et al (2020) Heterogeneity of treatment effect of prophylactic pantoprazole in adult ICU patients: a post hoc analysis of the SUP-ICU trial. Intensive Care Med 46:717–726

66. Ryan EG, Harrison EM, Pearse RM, Gates S (2019) Perioperative haemodynamic therapy for major gastrointestinal surgery: the effect of a Bayesian approach to interpreting the findings of a randomised controlled trial. BMJ Open 9:e024256

67. Laterre PF, Berry SM, Blemings A et al (2019) Effect of selepressin vs placebo on ventilator- and vasopressor-free days in patients with septic shock: the SEPSIS-ACT randomized clinical trial. JAMA 322:1476–1485

68. Young PJ, Nickson CP, Perner A (2020) When should clinicians act on non-statistically significant results from clinical trials? JAMA 323:2256–2257

69. Dahabreh IJ, Hayward R, Kent DM (2016) Using group data to treat individuals: Understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. Int J Epidemiol 45:2184–2193

70. Iwashyna TJ, Burke JF, Sussman JB, Prescott HC, Hayward RA, Angus DC (2015) Implications of heterogeneity of treatment effect for reporting and analysis of randomized trials in critical care. Am J Respir Crit Care Med 192:1045–1051

71. Caironi P, Tognoni G, Masson S et al (2014) Albumin replacement in patients with severe sepsis or septic shock. N Engl J Med 370:1412–1421

72. Mazer CD, Whitlock RP, Fergusson DA et al (2017) Restrictive or liberal red-cell transfusion for cardiac surgery. N Engl J Med 377:2133–2144

73. The SAFE Study Investigators (2004) A comparison of albumin and saline for fluid resuscitation in the intensive care unit. N Engl J Med 350:2247–2256

74. The PEPTIC Investigators for the Australian and New Zealand Intensive Care Society Clinical Trials Group, Alberta Health Services Critical Care Strategic Clinical Network, and the Irish Critical Care Trials Group (2020) Effect of stress ulcer prophylaxis with proton pump inhibitors vs histamine-2 receptor blockers on in-hospital mortality among ICU patients receiving invasive mechanical ventilation: the PEPTIC randomized clinical trial. JAMA 323:616–626

75. Andrews J, Guyatt G, Oxman AD et al (2013) GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol 66:719–725

76. Andrews JC, Schünemann HJ, Oxman AD et al (2013) GRADE guidelines: 15. Going from evidence to recommendation - determinants of a recommendation's direction and strength. J Clin Epidemiol 66:726–735

77. Kent DM, van Klaveren D, Paulus JK et al (2020) The predictive approaches to treatment effect heterogeneity (PATH) statement: explanation and elaboration. Ann Intern Med 172:W1–W25

78. Girbes ARJ, De GH (2020) Time to stop randomized and large pragmatic trials for intensive care medicine syndromes: the case of sepsis and acute respiratory distress syndrome. J Thorac Dis 12(Suppl 1):S101–S109

79. Burke JF, Sussman JB, Kent DM, Hayward RA (2015) Three simple rules to ensure reasonably credible subgroup analyses. BMJ 351:h5651

80. Altman DG, Royston P (2006) The cost of dichotomising continuous variables. BMJ 332:1080

81. Shehabi Y, Neto AS, Howe BD et al (2021) Early sedation with dexmedetomidine in ventilated critically ill patients and heterogeneity of treatment effect in the SPICE III randomised controlled trial. Intensive Care Med 47:455–466

82. Seymour CW, Kennedy JN, Wang S et al (2019) Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. JAMA 321:2003–2017

83. Adaptive Platform Trials Coalition (2019) Adaptive platform trials: definition, design, conduct and reporting considerations. Nat Rev Drug Discov 18:797–807

84. Simon N, Simon R (2013) Adaptive enrichment designs for clinical trials. Biostatistics 14:613–625

85. Pallmann P, Bedding AW, Choodari-Oskooei B et al (2018) Adaptive designs in clinical trials: Why use them, and how to run and report them. BMC Med 16:29

86. Talisa VB, Yende S, Seymour CW, Angus DC (2018) Arguing for Adaptive Clinical Trials in Sepsis. Front Immunol 9:1502

87. van Werkhoven CH, Harbarth S, Bonten MJM (2019) Adaptive designs in clinical trials in critically ill patients: principles, advantages and pitfalls. Intensive Care Med 45:678–682

88. Ryan EG, Bruce J, Metcalfe AJ et al (2019) Using Bayesian adaptive designs to improve phase III trials: a respiratory care example. BMC Med Res Methodol 19:99

89. Maia IS et al (2020) Antiviral agents against COVID-19 infection (REVOLUTION). https://clinicaltrials.gov/ct2/show/NCT04468087. Accessed 23 Sep 2021

90. Viele K, Broglio K, McGlothlin A, Saville BR (2020) Comparison of methods for control allocation in multiple arm studies using response adaptive randomization. Clin Trials 17:52–60

91. Thall PF, Fox P, Wathen J (2015) Statistical controversies in clinical research: Scientific and ethical problems with adaptive randomization in comparative clinical trials. Ann Oncol 26:1621–1628

92. Wathen JK, Thall PF (2017) A simulation study of outcome adaptive randomization in multi-arm clinical trials. Clin Trials 14:432–440

93. Berry SM, Connor JT, Lewis RJ (2015) The platform trial: an efficient strategy for evaluating multiple treatments. JAMA 313:1619–1620

94. Park JJH, Harari O, Dron L, Lester RT, Thorlund K, Mills EJ (2020) An overview of platform trials with a checklist for clinical readers. J Clin Epidemiol 125:1–8

95. Saville BR, Berry SM (2016) Efficiencies of platform clinical trials: a vision of the future. Clin Trials 13:358–366

96. Angus DC (2020) Optimizing the trade-off between learning and doing in a pandemic. JAMA 323:1895–1896

97. The REMAP-CAP Investigators (2021) Interleukin-6 receptor antagonists in critically ill patients with Covid-19. N Engl J Med 384:1491–1502

98. RECOVERY Collaborative Group (2021) Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. Lancet 397:1637–1645

99. The RECOVERY Collaborative Group (2021) Dexamethasone in Hospitalized Patients with Covid-19. N Engl J Med 384:693–704

100. The REMAP-CAP, ACTIV-4a, and ATTACC Investigators (2021) Therapeutic anticoagulation with heparin in critically ill patients with covid-19. N Engl J Med 385:777–789

101. Dimairo M, Pallmann P, Wason J et al (2020) The adaptive designs CONSORT extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. BMJ 369:m115

102. Ryan EG, Brock K, Gates S, Slade D (2020) Do we need to adjust for interim analyses in a Bayesian adaptive trial design? BMC Med Res Methodol 20:150

103. Dodd LE, Freidlin B, Korn EL (2021) Platform trials: beware the noncomparable control group. N Engl J Med 384:1572–1573

104. Schiavone F, Bathia R, Letchemanan K et al (2019) This is a platform alteration: a trial management perspective on the operational aspects of adaptive platform trials. Trials 20:264

105. Morrell L, Hordern J, Brown L et al (2019) Mind the gap? The platform trial as a working environment. Trials 20:297