

# BMJ Open Are large randomised controlled trials in severe sepsis and septic shock statistically disadvantaged by repeated inadvertent underestimates of required sample size?

Joshua L C Wong,<sup>1,2</sup> Alexina J Mason,<sup>3</sup> Anthony C Gordon,<sup>1</sup> Stephen J Brett<sup>1</sup>

**To cite:** Wong JLC, Mason AJ, Gordon AC, *et al*. Are large randomised controlled trials in severe sepsis and septic shock statistically disadvantaged by repeated inadvertent underestimates of required sample size? *BMJ Open* 2018;**8**:e020068. doi:10.1136/bmjopen-2017-020068

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-020068>).

Received 1 November 2017  
Revised 6 June 2018  
Accepted 8 June 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Department of Surgery and Cancer, Section of Anaesthetics, Pain Medicine and Intensive Care, Imperial College London, London, UK

<sup>2</sup>Department of Life Sciences, MRC Centre for Molecular Bacteriology and Infection, Imperial College London, London, UK

<sup>3</sup>Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK

## Correspondence to

Dr Joshua L C Wong;  
[joshua.wong@imperial.ac.uk](mailto:joshua.wong@imperial.ac.uk)

## ABSTRACT

**Objectives** We sought to understand why randomised controlled trials in septic shock have failed to demonstrate effectiveness in the face of improving overall outcomes for patients and seemingly promising results of early phase trials of interventions.

**Design** We performed a retrospective analysis of large critical care trials of severe sepsis and septic shock. Data were collected from the primary trial manuscripts, prepublished statistical plans or by direct communication with corresponding authors.

**Setting** Critical care randomised control trials in severe sepsis and septic shock.

**Participants** 14 619 patients randomised in 13 trials published between 2005 and 2015, enrolling greater than 500 patients and powered to a primary outcome of mortality.

**Intervention** Multiple interventions including the evaluation of treatment strategies and novel therapeutics.

**Primary and secondary outcome measures** Our primary outcome measure was the difference between the anticipated and actual control arm mortality. Secondary analysis examined the actual effect size and the anticipated effect size employed in sample size calculation.

**Results** In this post hoc analysis of 13 trials with 14 619 patients randomised, we highlight a global tendency to overestimate control arm mortality in estimating sample size (absolute difference 9.8%, 95% CI –14.7% to –5.0%,  $p < 0.001$ ). When we compared anticipated and actual effect size of a treatment, there was also a substantial overestimation in proposed values (absolute difference 7.4%, 95% CI –9.0% to –5.8%,  $p < 0.0001$ ).

**Conclusions** An interpretation of our results is that trials are consistently underpowered in the planning phase by employing erroneous variables to calculate a satisfactory sample size. Our analysis cannot establish if, given a larger sample size, a trial would have had a positive result. It is disappointing so many promising phase II results have not translated into durable phase III outcomes. It is possible that our current framework has biased us towards discounting potentially life-saving treatments.

## Strengths and limitations of this study

- This study captures large contemporary trials in severe sepsis and septic shock powered to mortality.
- Includes critical care trials independent of intervention employed.
- Examines evidence from sources outside of published trial manuscripts.
- Applies statistical analysis to the discrepancy between anticipated and actual variable used in trial design.
- The retrospective examination of effect size is limited by the proposed errors in sample size calculation.

## INTRODUCTION

The mortality from severe sepsis and septic shock has fallen demonstrably over the last few years.<sup>1</sup> Despite this, many large interventional trials in critical care, encompassing both novel therapeutics and optimised treatment strategies, have failed to confirm improved effectiveness of interventions in comparison with placebo or best routine care. Occasionally, the promise hinted at by small phase II efficacy trials has not been confirmed in larger phase III effectiveness trials.

During the planning of an interventional trial, investigators must determine both the population and the number of subjects to be enrolled. This is to ensure that the sample size is adequate to identify with reasonable statistical certainty both true differences and true lack of differences between groups, that is, avoiding both false-positive and false-negative results. To calculate the sample size, investigators identify a significance level (commonly 0.05) and a ‘power’ (commonly 80% to 90%). The significance level, often referred to as the alpha, is the chance of concluding there is evidence of a difference when actually there is not and the difference found in the trial was

due to chance. Crudely, the power is the capacity of the trial to avoid erroneously reporting no difference when in fact there is a true difference between treatments. The input factors required to calculate the required sample size in a trial powered to a dichotomous outcome such as mortality are:

1. The control arm event rate.
2. The effect size.
3. The alpha level—to control for type 1 error (false positive).
4. And the power—to control for type 2 error (false negative).

Control arm event rates are commonly derived from prior trials, historical epidemiological studies or inception studies. Investigators draw inferences on outcome event rate from a population akin to their prospective study population (by employing similar inclusion and exclusion criteria). The effect size is defined as the difference in primary outcome rate between the control and intervention arms of a trial. Anticipated effect sizes are typically estimated from earlier phase 'efficacy trials'; these data provide a prediction of the magnitude of an intervention in the trial population to be enrolled.

Given the appreciable numbers of recent trials that have not demonstrated statistically significant effectiveness, we speculated that the assumptions which investigators were making in planning their trials may have inadvertently designed in weaknesses such that the final result of the trial did not necessarily reflect the veracity or otherwise of the hypothesis being tested. To put this another way, we speculated that the assumptions used in determining trial design were erroneous and have systematically eroded the capacity of such trials to demonstrate either a true positive effect of the treatment being investigated or to confidently exclude any such treatment effect.

Thus, we sought to investigate the role of these two important estimated variables on the relative risk reduction in primary outcome measure. To do this, we examined first the difference between anticipated outcome control arm event rate used during trial planning and the actual control arm event rate in the trial report, and second the difference between the anticipated effect size of novel treatment and that subsequently reported in the trial publications.

## METHODOLOGY

This study was performed as a post hoc analysis of published results of clinical trials in severe sepsis and septic shock.

A MEDLINE search was performed in August 2016 to identify appropriate trials by using the medical subject heading (MeSH) terms 'sepsis', 'septic shock' and 'randomised control trial' with publication dates limited to 2005 to 2015. The publication list was independently accessed by two investigators (JLCW and SJB) and trials filtered against an a priori defined list of inclusion criteria (figure 1).

Those trials not matching all inclusion criteria were rejected. The two lists were then cross-checked to ensure uniformity. The list was subsequently passed on to a third sepsis expert investigator (ACG) to ensure the sample set was representative of the prevailing peer-reviewed literature. Two further trials were suggested at this point but were subsequently rejected, as the sample sizes were too small to be included in the analysis. The minimum number of enrolled subjects in trials was set at 500 for trials to be included in the analysis. This was somewhat arbitrary, but was chosen so that large *efficacy* studies were not included in the analysis; our objective was to explore the impact of our variables of interest in *effectiveness* trials with a clinical outcome measure—mortality. The statistical analysis plans from the papers identified were systematically analysed to collect information around trial power, anticipated control arm mortality and anticipated effect size. Sources of additional information included the examination of pretrial published protocols or statistical analysis plans or where necessary direct contact with the corresponding authors of the primary or protocol publications. The latter was undertaken if there were concerns that there could be potential ambiguity from our own interpretation of the published reports. These data were collated using Microsoft Excel (Microsoft, Seattle, Washington, USA), with statistical analysis performed using GraphPad Prism V.6 (GraphPad Software, La Jolla, California USA) and R V.3.3.1.<sup>2</sup>

Data were initially checked for normality using the D'Augustino-Pearson Omnibus test and subsequently analysed using t-tests or non-parametric equivalents where appropriate.

Measures of uncertainty, specifically CIs for the actual control arm mortality, intervention arm mortality and effect size, were calculated in R from the data provided in the trial reports using standard formulae.<sup>3</sup> The differences between the actual and anticipated control arm mortality have been summarised using a random-effects meta-analysis to allow for heterogeneity between studies. The power curve (figure 5) was created in R using the standard power calculation formula for comparing two proportions.<sup>4</sup>

## Patient and public involvement

Patients and public were not involved in the creation of this manuscript as it examines previously published study data.

## RESULTS

### Trials included in the final analysis

An initial MEDLINE search identified 251 articles matching the initial search criteria. Of these, 236 trials were excluded, as they did not meet the prespecified inclusion criteria; this resulted in 16 trials (figure 1).

Three trials were subsequently removed from the final analysis despite meeting the initial inclusion criteria. These were VISEP,<sup>5</sup> APROCCHS<sup>6</sup> and ART123.<sup>7</sup> VISEP



**Figure 1** Trial selection screening and selection.

and APROCCHS were excluded as one trial was stopped early for safety reasons and in the other the investigational drug (drotrecogin alpha (activated)) was withdrawn during the period of the trial. The resulting publications from these two studies did not provide the necessary data for us to use to explore our research questions. ART123 was a phase IIb trial, which has now been restarted as a phase III trial (ClinicalTrials.gov NCT01598831); we were unable to acquire sufficient data to explore our research questions from the original publication or following correspondence with the primary author. The 13 trials included in the final analysis are summarised in [table 1](#).

They form the basis for our exploration of control arm mortality and effect size estimate.

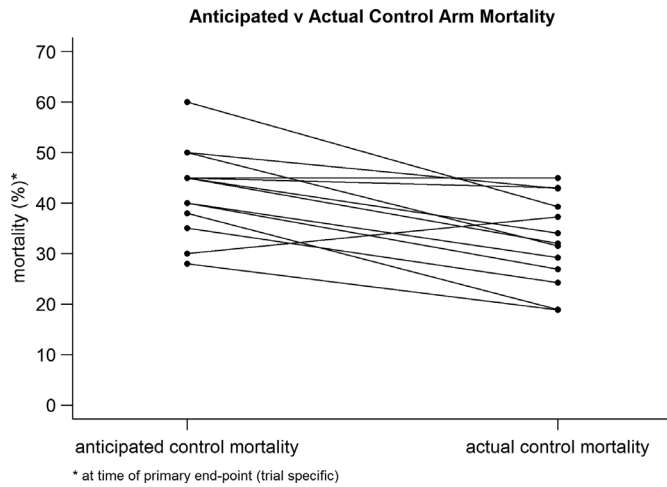
The primary end point of mortality was assessed at day 28 in seven trials, day 60 in one trial, day 90 in four trials and as in-hospital mortality in one trial. Sample size projections stated in statistical methods correlated well with the numbers of patients analysed in the final published intention-to-treat analysis groups. In ProCESS,<sup>8</sup> the sample size was recalculated after the first interim analysis, reducing the required sample size from 1950 to a total of 1341 patients at the end of enrolment. This recalculation preserved power at the same absolute effect

**Table 1** Trials data set included in final analysis

| Study                       | Publication year | Primary end-point mortality | Power (%) | Sample size calculation | Anticipated               |                           | Estimated absolute effect Size (%) | Control arm (n) | Intervention arm (n) | Control arm mortality (with 95% CI) | Intervention arm mortality (with 95% CI) | Control—intervention (with 95% CI) |
|-----------------------------|------------------|-----------------------------|-----------|-------------------------|---------------------------|---------------------------|------------------------------------|-----------------|----------------------|-------------------------------------|--|------------------------------------|
|                             |                  |                             |           |                         | control arm mortality (%) | control arm mortality (%) |                                    |                 |                      |                                     |  |                                    |
| SBITS <sup>18</sup>         | 2007             | 28                          | 90        | 800                     | 30                        | 10                        | 303                                | 321             | 37.3 (31.9 to 42.7)  | 39.3 (34.0 to 44.6)                 | -2.0 (-9.6 to 5.6)                       |                                    |
| VASST <sup>19</sup>         | 2008             | 28                          | 80        | 776                     | 60                        | 10                        | 382                                | 396             | 39.3 (34.4 to 44.2)  | 37.4 (32.6 to 42.2)                 | 1.9 (-4.9 to 8.7)                        |                                    |
| CORTICUS <sup>20</sup>      | 2008             | 28                          | 80        | 800                     | 50                        | 10                        | 248                                | 251             | 31.5 (25.7 to 37.3)  | 34.3 (28.4 to 40.2)                 | -2.8 (-11.0 to 5.4)                      |                                    |
| COIITS <sup>21</sup>        | 2010             | In-hospital                 | 80        | 508                     | 50                        | 12.5                      | 254                                | 255             | 42.9 (36.8 to 49.0)  | 45.9 (39.8 to 52.0)                 | -3.0 (-11.6 to 5.6)                      |                                    |
| PROWESS-SHOCK <sup>22</sup> | 2012             | 28                          | 80        | 1500*                   | 35                        | 7                         | 834                                | 846             | 24.2 (21.3 to 27.1)  | 26.4 (23.4 to 29.4)                 | -2.2 (-6.4 to 2.0)                       |                                    |
| 6S <sup>23</sup>            | 2012             | 90                          | 80        | 800                     | 45                        | 10                        | 400                                | 398             | 43.0 (38.1 to 47.9)  | 51.0 (46.1 to 55.9)                 | -8.0 (-14.9 to 1.1)                      |                                    |
| ACCESS <sup>24</sup>        | 2013             | 28                          | 90        | 2000                    | 40                        | 7.5                       | 657                                | 1304            | 26.9 (23.5 to 30.3)  | 28.1 (25.7 to 30.5)                 | -1.2 (-5.4 to 3.0)                       |                                    |
| TRISS <sup>25</sup>         | 2014             | 90                          | 80        | 1000                    | 45                        | 9                         | 502                                | 496             | 45.0 (40.6 to 49.4)  | 43.0 (38.6 to 47.4)                 | 2.0 (-4.2 to 8.2)                        |                                    |
| ProCESS <sup>26</sup>       | 2014             | 60                          | 80        | 1950*                   | 38                        | 6.5†                      | 456                                | 885             | 18.9 (15.3 to 22.5)  | 21.0 (18.3 to 23.7)                 | -2.1 (-6.6 to 2.4)                       |                                    |
| ARISE <sup>11</sup>         | 2014             | 90                          | 87.5†     | 1600                    | 28                        | 7.6                       | 804                                | 796             | 18.8 (16.1 to 21.5)  | 18.6 (15.9 to 21.3)                 | 0.2 (-3.6 to 4.0)                        |                                    |
| SEPSISPAM <sup>27</sup>     | 2014             | 28                          | 80        | 800                     | 45                        | 10                        | 388                                | 388             | 34.0 (29.3 to 38.7)  | 36.6 (31.8 to 41.4)                 | -2.6 (-9.3 to 4.1)                       |                                    |
| ALBIOS <sup>9</sup>         | 2014             | 28                          | 80        | 1350                    | 45                        | 7.5                       | 900                                | 895             | 32.0 (29.0 to 35.0)  | 31.8 (28.7 to 34.9)                 | 0.2 (-4.1 to 4.5)                        |                                    |
| ProMISE <sup>28</sup>       | 2015             | 90                          | 80        | 1260                    | 40                        | 8                         | 630                                | 630             | 29.2 (25.6 to 32.8)  | 29.5 (25.9 to 33.1)                 | -0.3 (-5.3 to 4.7)                       |                                    |

\*Recalculated on interim analysis.

†Calculated as the mid-point of quoted reference points.

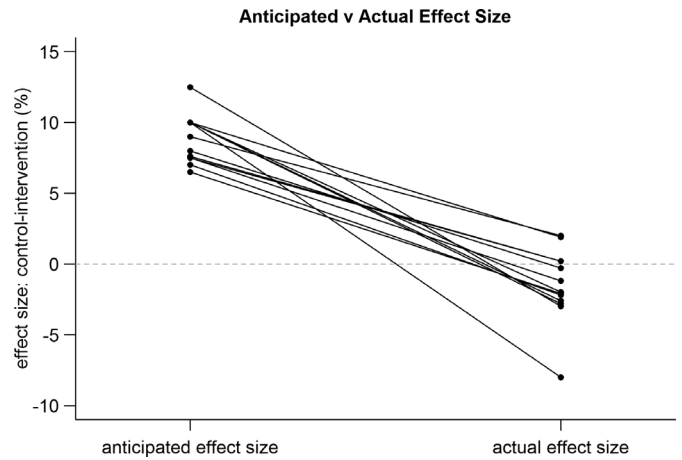


**Figure 2** Anticipated and actual control arm mortality. Comparison for individual trials.

size, and the investigators adjusted their analysis plan for the expenditure of power due to an interim analysis. The ALBIOS<sup>9</sup> trial size increased from an initial sample size of 1350 patients to a total of 1818 patients. This occurred after the second interim analysis by the Data Safety and Monitoring Board according to an a priori agreement with the investigators (L Gattinoni personal communication).

**Anticipated control arm mortality and effect size**

While all of the trials included patients with severe sepsis and septic shock, there remained considerable heterogeneity in detailed entry criteria. The trials and key characteristics are summarised in table 1. Figures 2 and 3 demonstrate the difference in anticipated and actual control arm mortality rate for the trials included in our study. Overall, there was a tendency to substantially

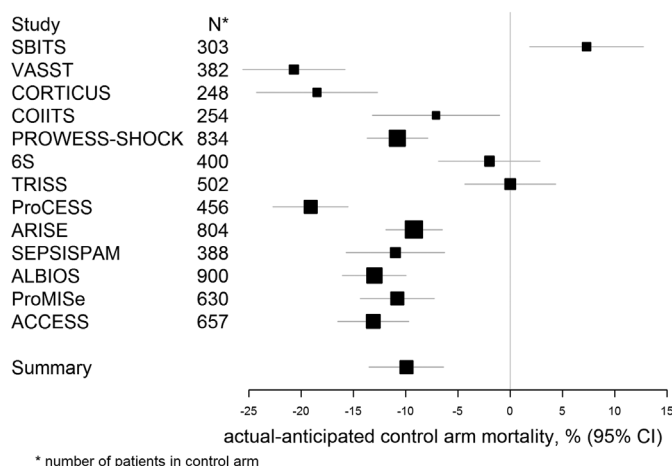


**Figure 4** Anticipated and actual effect size.

overestimate the control arm mortality, with strong evidence that the actual control arm mortality is lower than the anticipated control arm mortality (absolute difference, 9.8 percentage points; 95% CI -14.7% to -5.0%; p<0.001). In addition, there was a tendency to overestimate the treatment effect with very strong evidence that the actual effect size is smaller than the anticipated effect size (absolute difference, 7.4 percentage points; 95% CI -9.0% to -5.8%; p<0.0001), summarised in figure 4.

Taken overall, in these trials there has been a tendency to overestimate the anticipated rate of the primary outcome and overestimate the effect size of the treatment being investigated. This equates to a tendency for the trials to have an inadequate sample size resulting in less trial power than planned and increasing the risk of a type 2 error. The problem is demonstrated by the wide CIs around the estimates of mortality and effect size in table 1.

Difference between the Actual and Anticipated Control Arm Mortality



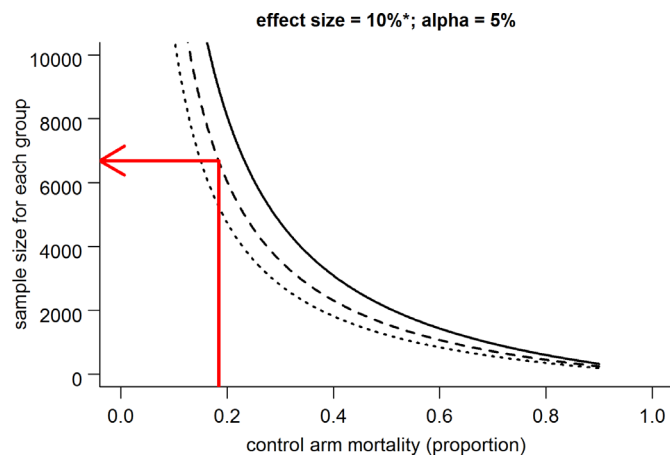
**Figure 3** Anticipated and actual control arm mortality. Forest plot of the results of a random-effects meta-analysis of the differences between the actual and anticipated control arm mortality (actual-anticipated). The horizontal lines correspond to the 95% CI for each study, with the corresponding solid square proportional to the weight for that individual study in the meta-analysis.

**DISCUSSION**

We set out to explore whether or not there was a consistent pattern of underestimating control group mortality rates and using overambitious estimates of treatment effect in the design of sepsis trials, and we chose to examine this using effectiveness trials in severe sepsis and septic shock. Our results suggest that both of these ideas are correct.

Despite the overall improvements observed in mortality,<sup>1</sup> the outlook for patients presenting with septic shock remains frustratingly uncertain. At best, an 18.8% control arm mortality rate (ARISE) still equates to 1 in 5 of this patient cohort dying as a result of their illness.

The consistent overestimation of control arm event rate (or lower-than-anticipated actual control arm event rate) may have systematically led to undersized trials from the outset, that is, given the actual control arm mortality the trials would have been designed to include more patients. This has likely meant that there has been an increased risk of type 2 errors in many sepsis trials that could have potentially resulted in the disregarding of potentially useful treatments.



**Figure 5** Power curve. The solid, dashed and dotted lines show the results of a standard sample size calculation to achieve 90%, 80% and 70% power, respectively. \*The effect size is specified as a 10% reduction relative to the control arm mortality, so in absolute terms the effect size gets smaller as the control arm mortality gets smaller. For example, if the control arm mortality is 18.4%, then this equates to an absolute reduction in mortality of 1.84%. The red arrow shows that for a 5% alpha, 6700 patients are required for each arm to achieve 80% power.

An anticipated control arm event rate will remain, with traditional prospective trial design, an estimate. Power calculations should take into account the uncertainty in this input factor. The source data on which the estimate will be based is always somewhat historical; investigators will attempt to match the proposed population of study with that of a similarly matched group of previous patients. The ‘goal posts’ are, however, moving. We know patients are increasingly doing better and while we have not been able to ascertain the exact factors that are driving these improvements, they have the potential to influence our ability to evaluate novel treatments. Even the ARISE inception study,<sup>10</sup> which collected data in 2006 and 2007 (although for only 3 months) and generated anticipated control arm data for the early goal-directed therapy (EGDT) trial ARISE<sup>11</sup> (enrolling in 2008 to 2014, n=1600), overestimated control arm mortality by 10% (28% vs 18.8%). There seem a number of plausible explanations. There are potential influences from the quality and assimilation of source data where investigators are not anticipating improvements in secular trends in outcome or investigators are using data from previous control arms and historical cohorts that are not sufficiently current. Furthermore, investigators may, in some cases, be bound, consciously or subconsciously, by economic constraints that limit the uppermost sample size that can feasibly be funded. The total number of subjects required in statistically evaluating an intervention, in the context of contemporary severe sepsis and septic shock mortality, is formidable. If we consider the sample size required in a trial that aims to demonstrate a relative effect size of 10% (with fixed variables of alpha (0.05) and power (70%, 80% or 90%)), then a fall in the population’s control arm

mortality has a profound effect (figure 5). This relative effect size applied to a contemporary mortality figure in septic shock of 18.4% would require a trial of 13 400 patients for the intervention arm’s mortality rate to fall to 16.56% (a 10% reduction relative to the control arm).

Effect size is a variable that is difficult to assess. We have used a post hoc analysis comparing the anticipated effect sizes that investigators estimated an intervention would have with the actual event rate in the study treatment arms. Investigators will use data from prior phase II (efficacy) studies, often with subtly differing enrolment and exclusion criteria from the subsequent phase III (effectiveness) trials. Phase II trials are conducted under as best circumstances as can be achieved and with a *per protocol* analysis of primary endpoint. Phase III effectiveness trials appropriately test interventions in the ‘real world’ and report using the *intention to treat* principle. This is a sterner/more complete test of an intervention. In addition, investigators estimate what they consider will be a clinically relevant effect.

The potential for patients enrolled in trials to fare better than those in a similar non-trial population is well recognised; teasing apart this ‘Hawthorne effect’ is difficult but may be important. The greater level of monitoring and general clinical surveillance may deliver subtle benefit. Trials of process of care may lose separation between groups, and thus erode measurement of treatment effect, as staff subtly alter their behaviour in response to what they see in the active or novel treatment arm. For trials aimed at delivering *overall* quality improvement, often with ‘cluster’ randomisation,<sup>12</sup> this is arguably not so important. For studies aimed at determining the absolute impact of particular interventions, including an ethnographic element to describe changes in care and some metrics of the actual ‘dose’ of the element of care under test delivered may be wise.<sup>13</sup>

In failing to demonstrate relatively large effects, these trials have not excluded small beneficial effects; however, the confident demonstration of smaller effects requires much larger trials, which may not be fundable, or the effects sufficiently persuasive to change practice. Further complexities include treatment heterogeneity across disease severity, and heterogeneity of treatment risk, which do not necessarily align; a hypothetical smaller beneficial effect may not be matched with a smaller side-effect risk.<sup>14</sup>

Using conservative estimates of event rate and effect size seems an obvious solution, however, in trial terms critical illness is ‘noisy’ and inflating numbers to overcome such noise in conventional trials is questionable. To quote David Sackett,<sup>15</sup> “Reducing confidence intervals by increasing the size of an RCT should be your last resort”.

In the absence of consensus opinion on an appropriate effect size that would sway clinicians to employ a new entity in clinical practice, where should we draw the line? Have we reached a point where traditional randomised control trial methodology no longer provides us with informative and practice changing results? Should clinicians leading

these trials, in collaboration with funders, shift towards composite endpoints (eg, long-term quality-of-life indices) instead of powering trials to a mortality benefit and would these metrics be sufficiently persuasive to change practice? Perhaps these difficulties in critical care research need to be addressed by exploring adaptive trial designs.<sup>16 17</sup> Based on a series of a priori determined decision rules and rolling interim analyses, such trials evolve and redesign themselves as they proceed. The challenge here may be communicating the theoretical benefits of such designs to funders and review boards, and not least learning how to explain these to potential study subjects—and their families.

**Contributors** JLCW: substantial contribution towards conception, design analysis and interpretation of the data. AJM: substantial contribution towards analysis and interpretation of data. ACG: substantial contribution in drafting work and revising it critically for important intellectual content. SJB: substantial contribution towards conception, design analysis and interpretation of the data. All authors approved of the final version. We agreed to be accountable for all aspects of the work and ensure accuracy and integrity.

**Funding** This study was supported by the National Institute for Health (NIHR) Comprehensive Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. ACG is funded by a National Institute of Health Research (NIHR) Research Professorship award (RP-2015-06-018)

**Disclaimer** The views expressed are those of the authors and not necessarily those of the NIHR, the NHS or the UK Department of Health.

**Competing interests** None declared.

**Patient consent** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** The data in this paper are available from the published peer-reviewed literature.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

- Kaukonen KM, Bailey M, Suzuki S, *et al*. Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000–2012. *JAMA* 2014;311:1308–9.
- R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2016.
- Kirkwood BR, Sterne JAC. *Essential medical statistics*. 2nd ed. New Jersey: Blackwell Publishing, 2003.
- Friedman LM, Furberg CD, DeMets DL. *Fundamentals of Clinical Trials*. 4 ed. New York, NY: Springer-Verlag New York, 2010.
- Brunkhorst FM, Engel C, Bloos F, *et al*. Intensive insulin therapy and pentastarch resuscitation in severe sepsis. *N Engl J Med* 2008;358:125–39.
- Annane D, Timsit JF, Megarbane B, *et al*. Recombinant human activated protein C for adults with septic shock: a randomized controlled trial. *Am J Respir Crit Care Med* 2013;187:1091–7.
- Vincent JL, Ramesh MK, Ernest D, *et al*. A randomized, double-blind, placebo-controlled, phase 2b study to evaluate the safety and efficacy of recombinant human soluble thrombomodulin, ART-123, in patients with sepsis and suspected disseminated intravascular coagulation. *Crit Care Med* 2013;41:2069–79.
- Yealy DM, Kellum JA, Huang DT, *et al*. A randomized trial of protocol-based care for early septic shock. *N Engl J Med* 2014;370:1683–93.
- Caironi P, Tognoni G, Masson S, *et al*. Albumin replacement in patients with severe sepsis or septic shock. *N Engl J Med* 2014;370:1412–21.
- Peake SL, Bailey M, Bellomo R, *et al*. Australasian resuscitation of sepsis evaluation (ARISE): a multi-centre, prospective, inception cohort study. *Resuscitation* 2009;80:811–8.
- Peake SL, Delaney A, Bailey M, *et al*. Goal-directed resuscitation for patients with early septic shock. *N Engl J Med* 2014;371:1496–506.
- Pearse R. Enhanced peri-operative care for high-risk patients (EPOCH) trial: a stepped wedge cluster randomised trial of a quality improvement intervention for .... *Lancet* 2015.
- Walsh TS, Salisbury LG, Merriweather JL, *et al*. Increased hospital-based physical rehabilitation and information provision after intensive care unit discharge: The RECOVER Randomized Clinical Trial. *JAMA Intern Med* 2015;175:901–10.
- Iwashyna TJ, Burke JF, Sussman JB, *et al*. Implications of heterogeneity of treatment effect for reporting and analysis of randomized trials in critical care. *Am J Respir Crit Care Med* 2015;192:1045–51.
- Sackett DL. Why randomized controlled trials fail but needn't: 2. Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!). *CMAJ* 2001;165:1226–37.
- Opal SM, Dellinger RP, Vincent JL, *et al*. The next generation of sepsis clinical trial designs: what is next after the demise of recombinant human activated protein C?\*. *Crit Care Med* 2014;42:1714–21.
- Bhatt DL, Mehta C. Adaptive designs for clinical trials. *N Engl J Med* 2016;375:65–74.
- Werdan K, Pilz G, Bujdoso O, *et al*. Score-based immunoglobulin G therapy of patients with sepsis: the SBITS study. *Crit Care Med* 2007;35:2693–701.
- Russell JA, Walley KR, Singer J, *et al*. Vasopressin versus norepinephrine infusion in patients with septic shock. *N Engl J Med* 2008;358:877–87.
- Sprung CL, Annane D, Keh D, *et al*. Hydrocortisone therapy for patients with septic shock. *N Engl J Med* 2008;358:111–24.
- Annane D, Cariou A, Maxime V, *et al*. Corticosteroid treatment and intensive insulin therapy for septic shock in adults: a randomized controlled trial. *JAMA* 2010;303:341–8.
- Ranieri VM, Thompson BT, Barie PS, *et al*. Drotrecogin alfa (activated) in adults with septic shock. *N Engl J Med* 2012;366:2055–64.
- Perner A, Haase N, Guttormsen AB, *et al*. Hydroxyethyl starch 130/0.42 versus Ringer's acetate in severe sepsis. *N Engl J Med* 2012;367:124–34.
- Opal SM, Laterre PF, Francois B, *et al*. Effect of eritoran, an antagonist of MD2-TLR4, on mortality in patients with severe sepsis: the ACCESS randomized trial. *JAMA* 2013;309:1154–62.
- Holst LB, Haase N, Wetterslev J, *et al*. Lower versus higher hemoglobin threshold for transfusion in septic shock. *N Engl J Med* 2014;371:1381–91.
- Yealy DM, Kellum JA, Huang DT, *et al*. A randomized trial of protocol-based care for early septic shock. *N Engl J Med* 2014;370:1683–93.
- Asfar P, Meziani F, Hamel JF, *et al*. High versus low blood-pressure target in patients with septic shock. *N Engl J Med* 2014;370:1583–93.
- Mouncey PR, Osborn TM, Power GS, *et al*. Trial of early, goal-directed resuscitation for septic shock. *N Engl J Med* 2015;372:1301–11.