Why we must cluster and cross over

Rinaldo Bellomo, Andrew Forbes, Muhammad Akram, Michael Bailey, David V Pilcher and D Jamie Cooper

The past decade has seen a marked increase in the number of Phase III[1-7] and, more recently, Phase II randomised controlled trials (RCTs)[8-11] conducted or designed by the Australian and New Zealand Intensive Care Society Clinical Trials Group (ANZICS CTG). Yet the identification of interventions that improve outcome has proved elusive. There are many significant challenges in identifying whether an intervention applied to a population of critically ill patients has an effect on mortality or any other patient-centred outcome measure (eg, the need for or duration of mechanical ventilation, the need for or duration of renal replacement therapy, disability-free survival or a favourable neurological outcome).

These challenges arise from the heterogeneity of intensive care unit patients, the fact that subpopulations of interest are created by the similarity of physiological states or syndromes rather than of disease states, and the myriad potentially confounding interventions that such patients are simultaneously exposed to. They also arise from the frequent monitoring of patients that allows countermeasures to be taken with high frequency, from our limited understanding of determinants of outcome and of the pathophysiology of the conditions we treat, and from the limited "strength" of our interventions.

Given such limitations, it is not surprising that a systematic analysis of major trials in intensive care has shown that investigators often hypothesise unrealistic effects when powering their trials. On average, the mean effect reported across such trials is a < 1.5% absolute risk reduction (ARR) in the outcome of interest.[12] If this is true, then the size of trials we would need to conduct to find such differences would involve the randomisation of > 10 000–15 000 patients. Such trials would be expensive and difficult to fund. Moreover, unlike conditions such as myocardial infarction, in which large numbers of patients are relatively easily identified, syndromes that occur in the ICU affect fewer patients and the accrual of such large numbers would be extraordinarily challenging.

If, for example, patients with sepsis were targeted for such interventions, in a routinely collaborative funding jurisdiction such as Australia and New Zealand, a trial of such magnitude would require $8–10 million and 6–8 years of recruitment time. As no studies in Australia or New Zealand have ever received more than $5 million in funding, and as a time frame of 6–8 years would likely deliver results that would have lost global relevance due to the evolution of critical care medicine, such a trial would simply never happen. Faced with these logistical challenges, investigators may choose to simply give up or continue to inflate the impact of the intervention and thus continue to deliver negative trials. This is clearly undesirable.

An alternative approach is to devise other strategies or trial designs to overcome these logistical and power difficulties. One possible approach is to stop randomly allocating patients and start randomly allocating ICUs.[13,14] Such cluster randomisation has several theoretical advantages over individual randomisation. It would occur in a more realistic setting, which would reproduce the way an ICU would function if the intervention was found to be effective and then applied to practice. It would allow recruitment of every patient of interest as an automatic default process that would markedly increase efficiency and numbers. It may simplify the consent process, as all patients (or all patients of interest in that unit) would receive the intervention, and an opt-out system could thus be applied. It may facilitate data collection as, in most participating units, such data collection might already be occurring. It might even decrease the cost of the trial because of the decreased burden of data collection.

However, in such cluster-based trials, the major determinants of statistical power are the number of clusters (in this case the number of ICUs) and the variation in outcome between clusters (operationalised by the so-called intracluster correlation coefficient).[15] The number of patients in each cluster has far less impact on statistical power. Given these characteristics, for a trial aimed at detecting a 2% ARR in the outcome of interest, a realistic number of clusters in typical ICU trials might be more than 200. This would make a cluster RCT impossible in Australia or New Zealand and overwhelmingly difficult globally.

A potential response to these challenges might come from the process of crossing over of clusters.[16] According to this alternative approach, clusters could be randomly allocated to receive treatment A or treatment B, then, after a suitable period (eg, 6 months or 1 year) swap to the alternative treatment, such that units randomly allocated to treatment A would then receive treatment B and those initially randomly allocated to treatment B would swap over to treatment A. This cluster cross-over (CXO) RCT design is significantly more powerful than the conventional parallel group cluster RCT, because comparisons are made within the cluster, thereby removing the variation between clusters that can confound conventional cluster trials, and leaving

only the variation in *changes* in outcome over time between clusters (eg, how clusters vary in their *changes* in mortality rates over time). When there is little such variation, the CXO design can be remarkably powerful and negate a large proportion of the loss of power due to the clustering in conventional parallel-cluster trials.

In the Australian and New Zealand jurisdictions, this regain of power can be estimated by interrogating the national database, which we have done. We have done the power calculations for a possible Australian and New Zealand CXO RCT in which the intervention would affect mortality. We found that a study using 35 Australian and New Zealand ICUs, with an average of 1000 patients per unit per observation period, could deliver a > 80% power to detect about 1.5% ARR in mortality. This number of clusters is relatively insensitive to the size of the clusters (number of patients admitted to the ICU or number of patients belonging to the subpopulation of interest). For example, reducing the number of patients to 500 per unit per period increases the number of units required to about 40. This is only a 14% increase relative to a halving of the number of patients per unit. There can also be a small loss of power due to variation in the throughput of patients across units per observation period, or of the same unit across periods.

In Australia and New Zealand this impact was slight and resulted in only a 10% increase in the number of units required. In terms of outcome event rate, as with individually randomised trials, for a given ARR, the closer the outcome rate is to 50%, the lower the power. However, the key issue under discussion here is the regain in power for a CXO design relative to a parallel-cluster design for a given research question and outcome measure, and here, the relative benefits of crossing over are fairly insensitive to the outcome rate.

This insight, derived from the evolution of trial design technology, has important implications, because even the current large studies (of 4000–5000 patients) being performed by the ANZICS CTG are still powered to detect ARRs in mortality of 4%–5%, a value threefold greater than could be detected by a CXO RCT. In addition, CXO RCTs have the major advantage of having a defined duration which is not affected by recruitment rates. This is likely to translate to faster trial turnover and a greater ability to address novel questions.

It is important to realise that CXO RCTs are less attractive for high-risk interventions when informed consent is crucial and a major ethical priority. They are, instead, ideal for very low-risk interventions in which both interventions being compared are within current practice, are administered according to clinician preference, and the difference in outcome between interventions is likely to be small. Examples include the use of saline or balanced solutions for fluid resuscitation and any crystalloid fluid therapy, or the comparison of two different caloric targets during enteral feeding. Such studies would compare accepted practices that might be delivered to ICU patients in different units according to clinician preference in any case. Such interventions could also be blinded to eliminate bias.

CXO RCTs for Phase IV interventions in critically ill patients that test the effectiveness of widely applied therapies provide a significant power advantage and important logistical and time advantages. They are likely to be much more cost-effective than current individual-patient randomisation trials. All the information increasingly available to us, as we explore more closely the statistical properties of such trial designs, indicates that this evolution is crucial to the trial agenda of the ANZICS CTG. The time to cluster and cross over is moving closer.

Competing interests

None declared.

Author details

Rinaldo Bellomo, Codirector[1]
Andrew Forbes, Professor of Biostatistics[2]
Muhammad Akram, Postdoctoral Fellow[2]
Michael Bailey, Associate Professor and Biostatistician[1]
David Pilcher, Director[3]
D Jamie Cooper, Codirector[1]

1 Australian and New Zealand Intensive Care Research Centre, Monash University, Melbourne, VIC, Australia.
2 School of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC, Australia.
3 Centre for Outcome and Resource Evaluation, Australian and New Zealand Intensive Care Society, Melbourne, VIC, Australia.

Correspondence: rinaldo.bellomo@austin.org.au

References

1 Finfer S, Bellomo R, Boyce N, et al; SAFE Study Investigators. A comparison of albumin and saline for fluid resuscitation in the intensive care unit. *N Engl J Med* 2004; 350: 2247-56.

2 NICE-SUGAR Study Investigators. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med* 2009; 360: 1283-97.

3 RENAL Replacement Therapy Study Investigators. Intensity of continuous renal-replacement therapy in critically ill patients. *New Engl J Med* 2009; 361: 1627-38.

4 Myburgh J, Li Q, Heritier S, et al. Statistical analysis plan for the Crystalloid versus Hydoxyethyl Starch Trial (CHEST). *Crit Care Resusc* 2012; 14: 44-52.

5 Myburgh JA, Finfer S, Bellomo R, et al. Hydroxyethyl starch or saline for fluid resuscitation in intensive care. *N Engl J Med* 2012; 367: 1901-11.

6 Aubron C, Carteaux G, Cooper DJ. Why we must "TRANSFUSE". *Crit Care Resusc* 2011; 13: 67-8.

7 Venkatesh B, Myburgh J, Finfer S. The ADRENAL study protocol: ADjunctive corticosteroid tREatment iN criticAlly ilL patients with septic shock. *Crit Care Resusc* 2013; 15: 83-8.

8 Young PJ, Saxena MK, Bellomo R, et al. The HEAT trial: a protocol for a multicentre randomised placebo-controlled trial of IV paracetamol in ICU patients with fever and infection. *Crit Care Resusc* 2012; 14: 290-6.
9 Dulhunty JM, Roberts JA, Davis JS, et al. The BLING II study: protocol for a multicentre randomised controlled trial of continuous beta-lactam infusion compared with intermittent beta-lactam dosing in critically ill patients with severe sepsis. *Crit Care Resusc* 2013; 15: 179-185.
10 Young P, Hodgson C, Dulhunty J, et al. End points for phase II trials in intensive care: recommendations from the Australian and New Zealand Clinical Trials Group consensus panel meeting. *Crit Care Resusc* 2012; 14: 211-5.
11 Myles PS, Bellomo R. A pivotal trial of fluid therapy for major abdominal surgery: need and equipoise. *Crit Care Resusc* 2011; 13: 278-80.
12 Aberegg SK, Richards DR, O'Brien JM. Delta inflation: a bias in the design of randomized controlled trials in critical care medicine. *Crit Care* 2010; 14: R77.
13 Climo MW, Yokoe DS, Warren DK, et al. Effect of daily chlorhexidine bathing on hospital-acquired infection. *N Engl J Med* 2013; 368: 533-42.
14 Huang SS, Septimus E, Kleinman K, et al. Targeted universal decolonization to prevent ICU infection. *N Engl J Med* 2013; 368: 2255-65.
15 Eldridge S, Kerry S. Sample size calculations. In: A practical guide to cluster randomized trials in health services research. Chichester, UK: Wiley, 2012: 137-71.
16 Turner RM, White IR, Croudace T; PIP Study Group. Analysis of cluster randomized cross-over trial data: a comparison of methods. *Stat Med* 2007; 26: 274-89. ❑

CriticalCare and Resuscitation

Journal of the College of Intensive Care Medicine of Australia and New Zealand • Chief Editor: Rinaldo Bellomo

Copyright transfer statement

The copyright to this article is transferred to the *College of Intensive Care Medicine of Australia and New Zealand* effective if and when the article is accepted for publication. The copyright transfer covers the exclusive right to reproduce and distribute the article, including reprints, translations, photographic reproductions, microform, electronic form (offline, online) or any other reproductions of similar nature.

We the undersigned declare that the material presented in the attached paper has not been published previously (nor is it currently being assessed for publication) by any journal other than *Critical Care and Resuscitation*.

MANUSCRIPT TITLE

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

AUTHORS' NAMES (Print)

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

AUTHORS' SIGNATURES (and Date)

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .