Challenging orthodoxy in critical care trial design: physiological responsiveness

Scott Aberegg

Jordan Valley Medical Center, West Jordan, UT, USA
Correspondence to: Scott Aberegg. Jordan Valley Medical Center, West Jordan, UT, USA. Email: scottaberegg@gmail.com.

Research involving critically ill adults poses unique challenges in addition to the usual difficulties involved in conducting quality, replicable scientific research. The critical care research community has responded admirably to these challenges by rigorously conducting numerous large and often multicenter randomized controlled trials (RCTs) of putative therapies for critical illnesses. Yet in spite of two decades of work, few incontrovertibly efficacious therapies have resulted from this herculean effort (1). The reasons for this are unsettled, but several categories of problems have emerged. The first problem is that of "positive" trials that cannot be replicated (2-15). Since the clinical trial is in essence a diagnostic test of a hypothesis (16), these non-replicable studies represent "false positives." False positive trials are due to type I errors which are increased by selection of a conventional and lax statistical significance threshold (e.g., α=0.05) (17,18), bias in the study at any stage of design, conduct, analysis and reporting (19), and fraud (20). Recently, the center for open science collaboration demonstrated that the majority of 100 "positive" psychological research studies could not be replicated, suggesting a false positive rate of 63% in that field (21). The generalizability of this result to medicine is uncertain, but the problems of non-replicability and false positives are not.

The second problem plaguing critical care research is a spate of negative trials of what were thought to be promising therapies. One possible explanation is that these negative trials represent "true negatives" and the trialed therapies do not work, for myriad reasons: unknown or redundant causal pathways to the outcome of interest (22), multiplicity of effects of the active treatment (pleiotropic and "off-target" effects), time dependency of causal pathways (23), etc. A second possibility is that some of the therapies are efficacious but for an outcome that was not assigned as the primary outcome (24) or was not measured at all. A third possibility is that the negative trials represent "false negatives" (16). False negative trials can result from inadequate assigned study power (25), from subversion of power calculations by delta inflation (use of an overly optimistic effect size in sample size calculations) (26), from inadequate dosing of active treatment causing failure of separation (27), or from dilution of effective sample size by patients unlikely to benefit because of severity of illness (too high or too low) (28,29) or because of heterogeneity introduced by non-specific disease definitions (30-33).

In a recent article (34), Goligher *et al.* propose one possible solution to the specific problem of false negative trials due to dilution of effective sample sizes. They reason that, prior to enrollment in a trial testing a physiological intervention for ARDS such as different doses of PEEP, a "test dose" of PEEP could be applied to prospective enrollees to determine PEEP responsiveness which would be an inclusion criterion for enrolment in the trial. By excluding patients who do not respond to PEEP with an increase in P/F ratio by a pre-specified margin, this strategy may exclude patients such as those with milder lung injury who cannot benefit from PEEP or those with severe disease and little recruitable lung who may be harmed by it (35). If this reasoning is correct, many fewer patients would need to be screened and enrolled to satisfy sample size requirements for such a trial.

The authors of this article are to be commended for scrutinizing the orthodoxy of contemporary trial design in light of its frequent failings and for proposing a possible solution. As the authors point out, for their strategy to work, the effect of PEEP on the outcome measure for the test dose (P/F ratio increase) must be an accurate predictor of the effect of PEEP on the primary outcome of the trial. That is to say, this strategy involves multiple bets (as do

all trials in the assignment of enrollment criteria). If the predictive validity of the screening outcome for the primary outcome is imperfect, we will have excluded otherwise eligible patients with this procedure. Unanswered questions include whether informed consent will be required prior to test dose administration, and what other diseases besides ARDS will lend themselves to screening with physiological responsiveness. For example, in a trial of vasopressors for shock, would we be content to administer a test dose of levophed and exclude from enrolment patients who did not have an increase in mean arterial pressure of a given amount? Is it possible that physiological responsiveness is dynamic and that failure to respond at one time does not predict failure to respond over the course of the illness?

While screening with physiological responsiveness is seductive for its potential to reduce sample size requirements, it ignores and potentially perpetuates larger problems in our current paradigms for evidence generation in critical care medicine. The exclusion of patients unlikely to benefit provides justification for using a larger delta value in power calculations. As there is no precedent in modern critical care for a reduction in short term proportional absolute mortality of 8–11% as the authors propose in Table 1, these values are likely to be overly optimistic and thus to represent delta inflation (26). Indeed, an emerging trend in critical care trials is to enroll not fewer but more patients, increasing the statistical precision of the results which better allows clinicians to exclude clinically meaningful effects outside the resulting 95% confidence intervals (36,37). The use of mortality as a universal primary endpoint for two decades without any consistent success in RCTs should lead us to reevaluate the suitability of this metric for the achievement of our goals. Other suggested measures include QALYs and composite outcomes (24,38) that include chronic encumbrances of critical illness such as artificial nutrition, supplemental oxygen, renal replacement therapy, and need for devices to assist with walking as components. Measurement of patient-centered outcomes such as these may inform not only our choices of outcomes in future trials if we see "signal" in an individual outcome, but will also help us refocus our attention on survivors of critical illness.

We have learned much in the modern era of critical care research. Our negative results should not discourage us, especially in light of data that outcomes in practice are improving in spite of them (39). Progress and future successes will depend not only on persistence and perseverance, but also on our willingness to challenge existing paradigms and dogma. Goligher *et al*. have taken us one step farther in the direction of progress.

Acknowledgements

None.

Footnote

References

1. Oba Y, Salzman GA. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury. N Engl J Med 2000;343:813; author reply 813-4.
2. 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.
3. Preiser JC, Devos P, Ruiz-Santana S, et al. A prospective randomised multi-centre controlled trial on tight glucose control by intensive insulin therapy in adult intensive care units: the Glucontrol study. Intensive Care Med 2009;35:1738-48.
4. Van den Berghe G, Wilmer A, Hermans G, et al. Intensive insulin therapy in the medical ICU. N Engl J Med 2006;354:449-61.
5. van den Berghe G, Wouters P, Weekers F, et al. Intensive insulin therapy in critically ill patients. N Engl J Med 2001;345:1359-67.
6. Abraham E, Laterre PF, Garg R, et al. Drotrecogin alfa (activated) for adults with severe sepsis and a low risk of death. N Engl J Med 2005;353:1332-41.
7. Ranieri VM, Thompson BT, Barie PS, et al. Drotrecogin alfa (activated) in adults with septic shock. N Engl J Med 2012;366:2055-64.
8. Bernard GR, Vincent JL, Laterre PF, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. N Engl J Med 2001;344:699-709.
9. Hypothermia after Cardiac Arrest Study Group. Mild therapeutic hypothermia to improve the neurologic outcome after cardiac arrest. N Engl J Med 2002;346:549-56.
10. Bernard SA, Gray TW, Buist MD, et al. Treatment of

comatose survivors of out-of-hospital cardiac arrest with induced hypothermia. N Engl J Med 2002;346:557-63.

11. Moler FW, Silverstein FS, Holubkov R, et al. Therapeutic hypothermia after out-of-hospital cardiac arrest in children. N Engl J Med 2015;372:1898-908.

12. Mourvillier B, Tubach F, van de Beek D, et al. Induced hypothermia in severe bacterial meningitis: a randomized clinical trial. JAMA 2013;310:2174-83.

13. Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 2001;345:1368-77.

14. ARISE Investigators; ANZICS Clinical Trials Group, Peake SL, et al. Goal-directed resuscitation for patients with early septic shock. N Engl J Med 2014;371:1496-506.

15. 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.

16. Browner WS, Newman TB. Are all significant P values created equal? The analogy between diagnostic tests and clinical research. JAMA 1987;257:2459-63.

17. Halsey LG, Curran-Everett D, Vowler SL, et al. The fickle P value generates irreproducible results. Nat Methods 2015;12:179-85.

18. Curran-Everett D. Explorations in statistics: hypothesis tests and P values. Adv Physiol Educ 2009;33:81-6.

19. Ioannidis JP. Why most published research findings are false. PLoS Med 2005;2:e124.

20. Alberts B, Cicerone RJ, Fienberg SE, et al. SCIENTIFIC INTEGRITY. Self-correction in science at work. Science 2015;348:1420-2.

21. Open Science Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science. Science 2015;349:aac4716.

22. Kerry R, Eriksen TE, Lie SA, et al. Causation and evidence-based practice: an ontological review. J Eval Clin Pract 2012;18:1006-12.

23. Weil MH. Lessons learned from clinical trials on monoclonal anti-endotoxin antibody. Arch Intern Med 1994;154:1183.

24. Aberegg S. Quality-adjusted life years or composite outcomes? Am J Respir Crit Care Med 2013;188:622.

25. Freiman JA, Chalmers TC, Smith H Jr, et al. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. N Engl J Med 1978;299:690-4.

26. 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.

27. Steinbrook R. How best to ventilate? Trial design and patient safety in studies of the acute respiratory distress syndrome. N Engl J Med 2003;348:1393-401.

28. Burke JF, Hayward RA, Nelson JP, et al. Using internally developed risk models to assess heterogeneity in treatment effects in clinical trials. Circ Cardiovasc Qual Outcomes 2014;7:163-9.

29. Wang H, Boissel JP, Nony P. Revisiting the relationship between baseline risk and risk under treatment. Emerg Themes Epidemiol 2009;6:1.

30. Beesley SJ, Lanspa MJ. Why we need a new definition of sepsis. Ann Transl Med 2015;3:296.

31. ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, et al. Acute respiratory distress syndrome: the Berlin Definition. JAMA 2012;307:2526-33.

32. Fröhlich S, Murphy N, Boylan JF. ARDS: progress unlikely with non-biological definition. Br J Anaesth 2013;111:696-9.

33. Kaukonen KM, Bailey M, Pilcher D, et al. Systemic inflammatory response syndrome criteria in defining severe sepsis. N Engl J Med 2015;372:1629-38.

34. Goligher EC, Kavanagh BP, Rubenfeld GD, et al. Physiologic Responsiveness Should Guide Entry into Randomized Controlled Trials. Am J Respir Crit Care Med 2015;192:1416-9.

35. Gattinoni L, Caironi P, Cressoni M, et al. Lung recruitment in patients with the acute respiratory distress syndrome. N Engl J Med 2006;354:1775-86.

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

37. Montori VM, Kleinbart J, Newman TB, et al. Tips for learners of evidence-based medicine: 2. Measures of precision (confidence intervals). CMAJ 2004;171:611-5.

38. Ferguson ND, Scales DC, Pinto R, et al. Integrating mortality and morbidity outcomes: using quality-adjusted life years in critical care trials. Am J Respir Crit Care Med 2013;187:256-61.

39. 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;311:1308-16.