HHS Public Access

Author manuscript

Intensive Care Med. Author manuscript; available in PMC 2022 December 22.

Published in final edited form as:

Intensive Care Med. 2022 June ; 48(6): 750–752. doi:10.1007/s00134-022-06691-4.

Characterizing systematic challenges in sample size determination for sepsis trials

Alexandre Tran[1,2,3,*], **Shannon M. Fernando**[3,4], **Bram Rochwerg**[5,6], **Christopher W. Seymour**[7,8,9], **Deborah J. Cook**[5,6]

[1]Department of Surgery, University of Ottawa, Ottawa, Canada.

[2]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada.

[3]Division of Critical Care, Department of Medicine, University of Ottawa, Ottawa, Canada.

[4]Department of Emergency Medicine, University of Ottawa, Ottawa, Canada.

[5]Department of Medicine, McMaster University, Hamilton, Canada.

[6]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada.

[7]Department of Critical Care, University of Pittsburgh, Pittsburgh, USA.

[8]Department of Emergency Medicine, University of Pittsburgh, Pittsburgh, USA.

[9]Clinical Research, Investigation, and Systems Modeling of Acute Illness (CRISMA) Center, Pittsburgh, USA.

Dear Editor,

Randomized controlled trials (RCTs) are considered the highest level of evidence for comparing health interventions [1]. However, inferences from trial results depend on clinical assumptions made during sample size determination. The sample size calculation for a superiority trial with a binary outcome incorporates [1]: (a) expected event rate in the control group (baseline risk), (b) target difference by the intervention (absolute or relative risk reduction), and (c) desired type I (p-value) and type II error (power) [1]. However, many sample size calculations are based on implausible assumptions about baseline risk and risk reduction [2]. The target difference should be informed by existing literature and important to patients [1]. Furthermore, prognostic or predictive enrichment strategies can be employed to inform more precise estimates of baseline risk or risk reduction, respectively [3].

The Surviving Sepsis Campaign (SCC) Guidelines [4] highlight the importance of an evidence-based approach to early identification and management. Despite the evaluation

*Correspondence: aletran@toh.ca.

Conflicts of interest

DJC is supported by a Canada Research Chair in Critical Care Knowledge Translation. None of the other authors report any conflict of interest.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s00134-022-06691-4.

of many interventions to improve outcomes for septic patients, few have shown reproducible benefit in clinical trials [5]. To understand sample size methodology for sepsis trials, we conducted a systematic review of RCTs evaluating interventions to reduce mortality in adults with sepsis, published in the year 2000 or later. The detailed methodology and results are provided in the online supplement. We included 60 RCTs (57,201 patients), most commonly based in Europe (33%) and comparing a pharmacologic intervention to placebo (52%).

For sample size determination (Table 1), baseline mortality was over-estimated by a median of 8% (1–14%). Only 8% of trials used prognostic enrichment to inform expected mortality in the control group. Fewer than 10% of trials provided clinical justification for the target difference. The median expected absolute risk reduction was 13% (9–20%) whereas the observed was 0% (− 3% to 4%). Studies were terminated early for futility (17%), signal suggesting harm (6%) or inadequate recruitment (3%). We found that 63% were completed but were unable to demonstrate the target difference. We evaluated the impact of the observed control group mortality and observed risk reduction on a revised sample size requirement and determined the reasons for inability to demonstrate the target difference (Flow Diagram in Supplement). These included observed lack of benefit or signal for harm (65%), overestimation of target difference (18%) and insufficient sample size to adequately evaluate the target difference (5%). To account for imprecision, we provide Forest Plots (Supplement) to demonstrate differences between the expected (pooed risk ratio [RR] 0.75, 95% CI 0.72–0.78) and observed treatment effects (pooled RR 1.01, 95% CI 0.97–1.04).

Imprecise baseline risk estimates and inflated treatment effect estimates are common in sepsis trial design [2] and can result in either over- or underestimation of sample size (Example in Supplement). Trialists have an ethical obligation to minimize these limitations by means of a greater emphasis on clinical justification, avoidance of improbable target differences [1], and utilization of prognostic enrichment to inform baseline risk [3]. Methodologic rigour and realistic sample size calculations could help to ensure the launch and conduct of trials that are most likely to inform practice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Cook JA et al. (2018) DELTA(2) guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. BMJ 363:k3750 [PubMed: 30560792]
2. Scales DC, Rubenfeld GD (2005) Estimating sample size in critical care clinical trials. J Crit Care 20(1):6–11 [PubMed: 16015511]
3. Stanski NL, Wong HR (2020) Prognostic and predictive enrichment in sepsis. Nat Rev Nephrol 16(1):20–31 [PubMed: 31511662]
4. Evans L et al. (2021) Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med 47:1181–1247 [PubMed: 34599691]
5. Cohen J et al. (2015) Sepsis: a roadmap for future research. Lancet Infect Dis 15(5):581–614 [PubMed: 25932591]

Table 1

Sample size calculation and outcomes

Description	*N* (%)
Sample size	
Sample size (actual), median (Q1–Q3)	444 (209–807)
Prognostic enrichment to inform baseline mortality risk (yes)	5 (8%)
Clinical justification for target difference (yes)	4 (7%)
Control event rate (mortality)	
Control mortality (expected), median (Q1–Q3)	43% (35–50%)
Control mortality (actual), median (Q1–Q3)	35% (27–43%)
Control mortality expectation achieved	13 (22%)
Intervention target difference calculation	
Based on relative risk reduction	12 (19%)
Based on absolute risk reduction	42 (66%)
Did not specify	6 (9%)
Intervention absolute risk reduction (mortality)	
Absolute mortality reduction (targeted), median (Q1–Q3)	13% (9–20%)
Absolute mortality reduction (actual), median (Q1–Q3)	0% (−3% to 4%)
Absolute mortality reduction target achieved	2 (3%)
Study result	
Terminated early (futility)	11 (17%)
Terminated early (signal suggesting harm)	4 (6%)
Terminated early (inadequate recruitment)	2 (3%)
No statistically significant treatment effect (completed)	38 (63%)
Statistically significant treatment effect (completed)	5 (8%)
Reason for inability to demonstrate target treatment benefit	
Observed absolute risk reduction ≤ 0% OR Study terminated for signal suggesting harm	35 (65%)
Observed absolute risk reduction > 0% BUT target difference is overly optimistic	10 (18%)
Observed absolute risk reduction > 0% BUT sample size insufficient to adequately evaluate target difference	3 (5%)
Inability to determine reason (insufficient information)	6 (11%)
Inadequate recruitment	1 (2%)