Injury, Int. J. Care Injured 41S (2010) S20–S23

Contents lists available at ScienceDirect

Injury

journal homepage: www.elsevier.com/locate/injury

Challenging issues in randomised controlled trials

A.D. Nichol [a,b,*], M. Bailey [a], D.J. Cooper [a,b]

On behalf of the POLAR and EPO investigators[1]

[a] Australian and New Zealand Intensive Care - Research Centre (ANZIC-RC), Monash University, Melbourne, Victoria 3004, Australia
[b] Alfred Hospital, Melbourne, Victoria 3004, Australia

ARTICLE INFO

Keywords:
Randomised controlled trial
Intensive care
Research
Education

ABSTRACT

What this topic is about: Randomised controlled trials (RCTs) are the most rigorous way of determining whether a cause–effect relation exists between treatment and outcome and are an integral component in the hierarchy of evidence which guide current clinical practice. Whether ensuring the success of a RCT or interpreting the medical literature, it is important to understand the key components of RCT design to assess their quality and therefore the weight which should be attributed to its findings. This article will highlight some of these key components by using a number of ongoing trauma studies being co-ordinated by the Australian and New Zealand Intensive Care Research Centre, Monash University.
Common problems and challenges: The quality of many RCTs could be improved by avoiding some common pitfalls, such as (i) unclear hypotheses and multiple objectives, (ii) poor selection of endpoints, (iii) inappropriate subject selection criteria, (iv) non-clinically relevant or feasible treatment/intervention regimens, (v) inadequate randomisation, stratification, blinding, (vi) lack of stratification in small RCTs (vii) inadequate blinding of trials, (viii) insufficient sample size/power, (ix) failure to use intention to treat analysis and (x) failure to anticipate common practical problems encountered during the conduct of a RCT.
Tips for researchers: The RCTs most likely to be funded and/or be of high quality always address these issues.

Introduction

"What this topic is about"

A randomised controlled trial (RCT) is a type of study commonly used in testing the efficacy or effectiveness of therapies or interventions. A relatively recent paradigm shift which has resulted in the prominence of randomised controlled trials in today's medical literature was partly triggered by a study comparing published randomised controlled studies with those that used observational designs.[14] This study demonstrated that biases in patient selection weighted the outcome of historical controlled trials in favour of new therapies, a flaw not apparent in high quality RCTs. In addition, there are a number of additional aspects of RCTs which augment their methodological robustness and have contributed to their adoption by clinical trialists worldwide. Importantly, the 'randomised' in RCT describes the processes ensuring random allocation of patients to either the treatment or control group (a placebo group not receiving the therapy). This process aims to ensure that no systematic baseline differences between treatment or control groups occur in factors, known and unknown, that may bias outcome other than the studied intervention.

Evidence-based medicine ranks studies according to 'grades' of evidence on the basis of their design.[8] Higher ranking studies are assigned a greater weight and are used to form guidelines and practice recommendations.[8] RCTs are among the highest grade of evidence (with the highest reserved for meta-analyses of high grade RCTs) while the lowest grade is applied to descriptive studies and expert opinion.[5] Randomised controlled trials have therefore become an integral part of current clinical practice. Interventional studies are commonly described as either phase I (studies documenting the safety of the intervention and usually performed on healthy volunteers), phase II (the treatment/intervention is

* Corresponding author at: Australian and New Zealand Intensive Care Research Centre, School of Public Health and Preventive Medicine, Monash University, Alfred Hospital Campus, 89 Commercial Road, Melbourne 3004, Australia.
Tel.: +61 3 99030247; fax: +61 3 99030071.
E-mail address: alistair.nichol@med.monash.edu.au (A.D. Nichol).

[1] *The POLAR Study investigators*: D James Cooper, Stephen Bernard, Alistair Nichol, John Myburgh, Jeffery Presneil, Peter Cameron, Andrew Forbes, Lynne Murray, Jeffery Rosenfeld. *The EPO-TBI Study investigators*: Rinaldo Bellomo, D James Cooper, Alison Street, Alistair Nichol, Craig French, Jeffery Presneil, John McNeil, Neil Orford, Jeffery Rosenfeld, Gillian Syres, Michael Bailey, Ann Mak, Shirley Vallance, Lorraine Little.

0020–1383/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.injury.2010.03.033

given to a small group of patients, with endpoints frequently surrogate markers of improved clinical outcomes such as oxygenation, cytokine concentrations, etc.), phase III (effectiveness trials in larger groups of patients which are powered to detect clinically relevant endpoints, such as mortality) and phase IV (post marketing studies to monitor for adverse events or to expand the indication for a therapy). For the most part, phase II and phase III trials are randomised controlled trials.

While randomised controlled trials are useful in many settings it is also important to note that they may not be the most appropriate design for the evaluation of all treatments/interventions, in particular those that have rare outcomes or effects that take a very long time to develop. In such instances, other study designs such as case-control studies or cohort studies may be more appropriate. Although it should be noted that the RCT would still give you the most robust estimate of effect but on the basis of feasibility (not robustness) these other designs are frequently selected.

Whether working to design a future RCT or interpreting the medical literature it is important to understand the key components of a RCT to assess its quality and therefore the weight which should be given to its findings. This article will for illustrative purposes highlight some of these key components by using a number of currently funded traumatic brain injury (TBI) studies currently being conducted by the Australian and New Zealand Intensive Care Research Centre (ANZIC-RC), Monash University, Melbourne. In particular reference will be made to the POLAR (The Prophylactic Hypothermia Trial to Lessen Traumatic Brain Injury, NCT00987688) and EPO-TBI (Erythropoietin in Traumatic Brain Injury, NCT00987454) RCTs. While this review will briefly highlight the key components of a RCT, it should also be noted that a number of exhaustive high quality guidelines for reporting and critiquing RCTs have been published.[1,11,2]

"Common problems and challenges" and "Tips for researchers"

Frequently encountered problems and challenges with RCTs include

Unclear hypotheses and multiple objectives

It is vital to have clear objectives which can be rigorously tested. Not only does a clear hypothesis with a single primary objective allow adequate statistical planning and power analysis, it also states up front the clinical purpose of the trial and frames its clinical relevance. For example, the objective of the POLAR trial is to determine if 'early prophylactic hypothermia compared to normothermia improves early neurological outcomes 6 months after severe traumatic brain injury' and the EPO-TBI study aims to determine if 'erythropoietin 40,000 U compared to placebo weekly for 3 weeks will improve neurological outcomes 6 months following moderate or severe TBI'. The answers from these studies will either be 'yes it does' or 'no it doesn't' and the result will be useful in addressing an important and relevant clinical question and guiding practice in these areas.

Note: Keep it simple, and clear; too many objectives muddle a trial. A single hypothesis with only a few secondary hypotheses will suffice.

Poor selection of study endpoints

Endpoints need to be clinically relevant, there is no point reporting or testing a clinical endpoint which will not aid the clinician in deciding whether a particular treatment should be given to a patient or not. While phase II trials frequently utilise surrogate markers of outcome (physiological, biochemical or inflammatory), phase III studies generally test 'harder' markers of outcome such as mortality. There is an increasing recognition that there are more important things from a patient and societal point of view rather than just being dead or alive. Many phase III trials in TBI therefore use internationally recognised and validated tools to assess neurological function (i.e. Glasgow Outcomes Scale, Extended (GOSE)) and quality of life (i.e. EQ-5D). Therefore we have decided that both the POLAR and EPO-TBI trials will assess neurological function 6 months after injury (GOSE) as their primary endpoint and quality of life (EQ-5D) as a secondary endpoint.

Note: Pick the one primary variable that adequately measures the most important, relevant or practical outcome of interest.

Inappropriate subject selection criteria

The patient inclusion criteria in a RCT will describe the study population and therefore needs to clear, objective and clinically relevant, i.e. POLAR and EPO will enrol patients with either severe TBI (Glasgow Coma Score (GCS) <9) or severe and moderate TBI ($GCS < 13$), respectively. It is entirely reasonable to impose exclusion criteria to prevent recruitment of patients who may be at risk of harm by the treatment/intervention and to prevent enrolment of patients in whom treatment is futile (i.e. patients with fixed and dilated pupils and GCS = 3 in TBI studies). But, it is important that these exclusion criteria are not overly complex and do not significantly reduce the generalisability of the final study results. These exclusions may be general (i.e. pregnancy) or specific (i.e. induced hypothermia may be associated with coagulopathy therefore it would be prudent to exclude trauma patients at high risk of bleeding from enrolment in the POLAR trial).

Note: Balance the benefits and risks of selecting a very homogeneous patient group (possibly easier to detect effects but harder to enrol subjects and less generalisable results) versus heterogeneous patient selection criteria (easier to enrol subjects, more generalisable results but possible dilution of treatment effects).

Non-clinically relevant or feasible treatment/intervention regimens

Common sense would dictate that the most useful RCTs not only determine efficacy but that the study treatment/intervention can easy be implemented in the study population in most clinical settings. Sometime extremely complex or labour intensive treatments/interventions while potentially beneficial cannot be implemented due to logistical and resource issues. For these reasons we have simplified the study interventions in both the EPO-TBI and the POLAR trials to a degree that if efficacious, all Australian trauma centres would have little difficulty adopting these interventions.

Note: As a rule, choose study treatments/interventions which would be practical in most relevant clinical settings.

Inadequate randomisation, stratification, blinding

Random allocation means that participants will be randomly assigned to either the study or control group. Therefore, if the sample size is adequate the characteristics of the participants are likely to be similar between groups at the start of the study. If randomisation is conducted properly and in a large cohort of patients, it reduces the risk of serious imbalances in known and unknown factors that could falsely influence the endpoints being assessed. As random allocation gives the randomised controlled trial a significant ability to reduce bias, it is therefore of vital important that it is correctly performed and documented. There are various methods of randomisation with the common element in each being that no one should be able to determine ahead of time which group a given patient will be assigned to. Despite this being a central element of any RCT, randomisation method is poorly and infrequently reported.[15,17,10]

Ideally, the description of randomisation will describe, (i) the design of the randomisation schedule (i.e. computer based schedule with block sizes described), (ii) whether there is stratification or not, (iii) who randomizes the patient and how (i.e. bed side nurse and web protocol), (iv) record and log all randomisation attempts and describe a process by which confirmation of correct allocation to either the treatment or control group after randomisation has occurred.

Lack of stratification in small RCTs

Stratification is a method which aims to control for important prognostic factors that may be imbalanced between treatment arms. For example, the EPO-TBI study will stratify for severity of brain injury (moderate or severe) to ensure that equal numbers of patients with similar degrees of injury are represented in both groups. By balancing groups over the strata of interest, it increases the interpretability of the findings and frequently avoids the need for multivariate adjustment to be performed should baseline imbalances arise by chance. The use of stratification is particularly important for smaller trials with less than 200 patients where a lack of stratification could cause insufficient numbers for meaningful sub-group analysis.

Note: Stratify to prevent imbalances in important prognostic factors between groups especially if the study is small or the prognostic factor is very powerful.

Inadequate blinding of trials

Inadequate blinding can introduce bias because known treatment allocation may cause a patient to be treated differently in the broadest sense, i.e. followed up differently, investigated differently, not just given different therapies, thus influencing the outcome of the study. Numerous techniques can be used to maintain blinding and ensure that outcomes are objectively assessed such as the use of placebo treatment or sham procedure. For instance our EPO-TBI trial is a double blinded placebo controlled RCT where neither the treating physician nor the researcher assessing 6 month neurological outcomes will know treatment allocation (only a research pharmacist who will prepare identical EPO and placebo syringes). It is important to note that it is not always possible to blind treatment assignments (patients in the POLAR study will be maintained either hypothermic (33C) or normothermic (37C) for at least 3 days in ICU) but outcome evaluation can be blinded (in POLAR the trained outcome assessor will be blinded to allocation). This blinding is most important for studies whose endpoints are subjective (pain, function, etc.).

Note: As a minimum, always blind outcome assessor to treatment allocation.

Insufficient sample size/power

Obtaining statistically significant differences between two groups is relatively straightforward if large differences are expected. However, many randomised controlled trials need large sample sizes because the studied treatments are expected to have a relatively small effect. The sample size required to achieve power in a study is inversely proportional to the treatment effect squared.[16] Unfortunately a relatively common criticism of RCTs is that the sample size was inadequate to detect a clinically plausible difference between the study groups. This is nicely demonstrated by a recent study of 71 randomised controlled trials that showed most of these trials were too small (i.e. had insufficient power) to detect important clinical differences. More concerning however was the fact that the authors of these trials seemed unaware of these facts.[9] In addition, recent studies of patients enrolled in schizophrenia and TBI studies found that only a very small number of conducted trials were adequately powered.[17,7]

For example, in the POLAR study we used recent prospective studies[13,12] to determine a weighted mean rate of current favorable neurological outcomes of 50% in Australian and New Zealand patients with severe TBI. Furthermore a beneficial effect of prophylactic hypothermia in severe TBI of 30% relative risk increase (RRI) from 50% to 65% (15% absolute risk increase (ARI)) in favourable neurological outcome at 6 months following injury would be a clinically relevant and important effect. This was determined to be plausible and conservative estimate, based on:

- A recent meta-analysis[3] in severe traumatic brain injury comparing prophylactic hypothermia to normothermia which demonstrated a 46% improvement of favourable outcomes (Relative Risk (RR) of 1.46, 95% (CI) 1.12–1.92, $p = 0.006$).
- A finding of a 50% increase ($p = 0.02$) in favourable outcomes in a sub-group of patients with severe traumatic brain injury <45 years of age who were hypothermic on arrival and subsequently randomised to hypothermia versus normothermia.[4]

A 30% RRI in favourable neurological outcomes is approximately two-thirds of the RRI demonstrated elsewhere.[6] If prophylactic hypothermia was proven to be beneficial, such a difference in neurological outcomes would be highly clinically significant (NNT = 7) and would lead to a widespread change in management of severe TBI patients in Australia and internationally.

Note: Keep treatment effect size conservative. Error in this aspect of trial design is a common reason for funding failure.

Failure to use intention to treat analysis

It is important that the primary outcome analysis is undertaken in patients within the group to which they were allocated, irrespective of whether they experienced the intended intervention or not. This intention to treat analysis (ITT) analysis is intended to avoid various misleading biases that can arise and is based on the assumption that, as in real life, sometimes patients do not all receive optimal treatment, even though that was the initial intention. ITT provides information about the potential effects of TREATMENT POLICY (i.e. the decision to treat all) rather than on the potential effects of SPECIFIC TREATMENT (i.e. the effect on the group who receive the treatment as planned). While cross over and drop out in an ITT will still give you a biased estimate of the true effect of the intervention, the key is that this will always biased towards the null hypothesis and this give you a conservative result.

The main reason to do an ITT analysis is to avoid the potential bias that crossover and drop out can introduce, thus preserving the principles of randomisation. Using an example from the POLAR trial, if patients who had a more severe TBI and subsequent increased likelihood of death were unable to tolerate the 3 days of induced hypothermia; they developed complications and were re-warmed. At the end of the study if we conduced a per-protocol analysis, were only patients who received the full course of cooling were compared to controls, it may appear that hypothermia improved outcomes as we had systematically excluded the more severely injured patients in the hypothermia group in the analysis. An ITT analysis would take these patients into consideration. Therefore, everyone who begins the treatment is considered to be part of the trial!

Note: Always use intention to treat analysis

Practical problems encountered during the conduct of a RCT

Many problems can frustrate the successful completion of a trial once started. Therefore, the proactive researcher will attempt to anticipate these problems.

Low recruitment. This is the perennial problem with all RCTs which may arise from overly selective inclusion criteria, too many inappropriate exclusion criteria, inadequate assessment of the

number of patients available to be entered into the study, inadequate detection of potentially recruitable patients (screening) and not accounting for refusal of consent. Determine the case rate (with inclusion and exclusion criteria considered) and then at most, anticipate recruiting fewer than 50% of these potentially eligible patients. Both EPO and POLAR trials have anticipated recruitment rates of below 50% of the potential patients, as determined by a previous multi-centre observational study.[13]

Note: Be conservative only anticipate to recruit under 50% of the potentially eligible patients.

Crossover, withdrawal or loss to follow-up. Not all patients will remain in their allotted treatment group secondary to physician choice (although a protocol violation), the development of complications, withdrawal of consent, etc. It is important to attempt to anticipate these occurrences. In addition, there will always be a number of patients who it will be unable to assess the primary endpoint; this is more common if the endpoint is measured a significant time after the initial illness or after the patients has been discharged from hospital.

For example, in the POLAR study to ensure availability of the required number of evaluable patients at the end of the study, the sample size was inflated to account for losses to (i) follow-up, (ii) withdrawal of hypothermia, (iii) due to contraindications. Allowing an overall proportion of 5% loss to long term follow-up, expanded the sample size to 384 patients. Furthermore, based on pilot data, allowance for a 12% rate of withdrawal of the hypothermic intervention for patients randomised to hypothermia (8% bleeding, 2% withdrawal of surrogate/physician consent and an estimated 2% rate of inappropriate enrolment (i.e. cardiovascular accident not TBI) increased the sample size to 496 patients.

Note: Always anticipate some cross over and loss to follow up to adequate maintain study power.

Financial support

No funding was received for the preparation of this manuscript. Dr A Nichol is the recipient of a Victorian Neurotrauma Initiative (VNI) early career practitioners fellowship, Prof Cooper is the recipient of a National Health and Medical Research Council (NHMRC) practitioner fellowship.

References

1. Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. BMJ 1996;313:570–1.
2. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637–9.
3. Brain_Trauma_Foundation. Guidelines for the management of severe traumatic brain injury. J Neurotrauma 2007;24.
4. Clifton GL, Miller ER, Choi SC, et al. Lack of effect of induction of hypothermia after acute brain injury. N Engl J Med 2001;344:556–63.
5. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000;342:1887–92.
6. Cooper DJ, Myles PS, McDermott FT, et al. Prehospital hypertonic saline resuscitation of patients with hypotension and severe traumatic brain injury: a randomized controlled trial. J Am Med Assoc 2004;291:1350–7.
7. Dickinson K, Bunn F, Wentz R, Edwards P, Roberts I. Size and quality of randomised controlled trials in head injury: review of published studies. BMJ 2000;320:1308–11.
8. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992;268:2420–5.
9. Freiman JA, Chalmers TC, Smith Jr H, Kuebler RR. 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.
10. Jadad AR, Rennie D. The randomized controlled trial gets a middle-aged checkup. JAMA 1998;279:319–20.
11. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 2001;285:1987–91.
12. Myburgh J, Cooper DJ, Finfer S, et al. Saline or albumin for fluid resuscitation in patients with traumatic brain injury. N Engl J Med 2007;357:874–84.
13. Myburgh JA, Cooper DJ, Finfer SR, et al. Epidemiology and 12-month outcomes from traumatic brain injury in Australia and New Zealand. J Trauma 2008;64:854–62.
14. Sacks H, Chalmers TC, Smith Jr H. Randomized versus historical controls for clinical trials. Am J Med 1982;72:233–40.
15. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408–12.
16. Stolberg HO, Norman G, Trop I. Randomized controlled trials. Am J Roentgenol 2004;183:1539–44.
17. Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. BMJ 1998;317:1181–4.