B. Ollivere
Editor-in-Chief
editor360@boneandjoint.org.uk

Can we get more from large randomized controlled trials?

I cannot help but be honest about my enthusiasm for large randomized controlled trials (RCTs). It seems to me that the concept of testing an intervention by experiment, randomly assigning patients to one group or another, utilizing the same outcome measures, and blinding as many people as possible to the treatment allocation and assessing patients in the same way has an appeal that other forms of experiment do not.

The problem, of course, is that these large trials are of significant cost to the funder, are designed by their very nature to answer just a single question, and take a huge amount of resources to complete. When designing and setting up a trial, the preparatory work includes settling on the trial question (with suitable equipoise), working out the best way to measure an outcome, defining what a meaningful difference would be, and determining when to measure the outcome. The process of powering the primary outcome measure is driven by the decisions above, giving a sample size calculation usually to ensure a type I error of 5% or less (finding a false positive) and a type II error of 20% (finding a false negative).

This powering difference in terms of type I and type II errors perhaps partly explains why there are more 'negative' than 'positive' trials out there. However, it is important to remember that if trialists and clinicians get the questions exactly at equipoise, odds are that half of the trials will be positive and half negative. One of the real difficulties with large randomized trials is the reporting of secondary outcome measures. A 'hardcore' methodologist would likely hold the line that the primary outcome measure is the only one to look at, and that all secondary outcomes are hypothesis-generating. They may even go so far as to say that secondary outcomes are irrelevant to the message of the trial. Others will say that secondary outcomes are just as valid, and many smaller studies are reported without recourse to a specific primary outcome.

The correct position is clearly somewhere between these two arguments. Few would claim that complication rates, which are of course a secondary outcome, should be ignored. Prespecified subgroup analyses are often considered a reasonable approach, as they help to avoid the scattergun approach of multiple subgroup analyses. Whatever the rights and wrongs, clarity of reporting is essential. Presenting the results in a clear and unbiased way will lead to easier interpretation of the results.

To help us with this, there are a range of different standards; the Consolidated Standards of Reporting Trials (CONSORT) statement, which applies to RCTs only, is one of the most well-known. However, there are a wide range of standards for reporting studies, such as the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement for epidemiological studies, and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement for systematic review. While these standards are there to help with reporting, they often do not help with strengthening interpretation or methodology, and simply make it easier to spot when something is not quite right.

The Enhancing the Quality and Transparency of Health Research (EQUATOR) network attempts to improve reporting of all types of studies. EQUATOR define themselves as "an international initiative that seeks to improve the reliability and value of published health research literature by promoting transparent and accurate reporting and wider use of robust reporting guidelines". Even within the EQUATOR network, there is little to help tighten up the interpretation of trials. There are few hard-and-fast guidelines, use of assessment of bias tools, and high-quality evidence appraisal and synthesis such as the approaches offered by the Cochrane collaboration.

Despite all of this methodological rigour, I cannot help but feel we may be able to make more of these very rich datasets that are being collected as part of large multimillion-pound RCTs.

 DOI: 10.1302/2048-0105.86.360739