Surgical trial design – learning curve and surgeon volume

Determining whether inferior results are due to the procedure itself, or delivery of the procedure by the surgeon


The British Editorial Society of Joint Surgery, London, United Kingdom

A fundamental difference between drug trials and surgical trials is that the latter involve a procedure and therefore the patient outcome will depend on a complex interaction between surgeon, patient and the operating environment. With regards to the surgeon’s impact, the outcome will be influenced by their proficiency in the delivery of the procedure. Surgical proficiency is known to depend on both the learning curve for a procedure and the case volume of the surgeon, i.e., it is necessary to both ascend the learning curve and to practice the procedure regularly.

The learning curve varies for different procedures. For instance for hip fracture fixation the operating time does not level off until the surgeon has done 20 to 30 of these operations. For reduction of uncomplicated congenitally dislocated hips it may be slightly fewer as the performance seems to level after 10 to 20 procedures and can be accelerated with appropriate simulation training. For arthroscopy of the knee, a simulation study suggests that consultant level skills are not reached until 170 procedures have been carried out. Thus although the actual number might vary, there will always be a certain number of cases a surgeon needs to have performed before his/her proficiency has plateaued. From the ethical point of view, before offering a new procedure a surgeon should be confident that he/she will do sufficient procedures, the learning curve is part and parcel of that effectiveness — in the real world, the surgeons will have to ascend that learning curve on real patients, whose outcomes should count in the overall assessment of the two procedures.

In some trials the condition for which the procedure is carried out and the learning curve would be expected to have been completed during orthopaedic training e.g primary hip replacement, closure of knee replacement wounds. Whereas in other studies the procedure is more specialised and the condition less common (e.g. pedicle lengthening for spinal stenosis); a surgeon is less likely to complete the learning curve during training. In studies with low numbers, it is particularly important to ensure that the learning curve has been completed. A more difficult issue occurs when a randomised trial is comparing a new with an established surgical procedure, and the surgeons may well be on different points of the learning curve for the new and the established procedures. Then a crude comparison of the outcomes might be unfair to the new procedure, as it includes less experience with the new technique, and so incorporates more of the transient ascent to the plateau of that new learning curve. Some methodologists have developed the concept of expertise based trials to partially address the problem of learning curves, such as that reported in Devereaux et al. Another view is that although it might be methodologically correct to adjust out the learning curve in an efficacy-based randomised trial, if one is instead interested in the comparative effectiveness of two surgical procedures, the learning curve is part and parcel of that effectiveness — in the real world, the surgeons will have to ascend that learning curve on real patients, whose outcomes should count in the overall assessment of the two procedures.

Modelling of the learning curve has been discussed by several authors, and can help place the findings of a study in clinical context. However, it still remains that if clinical trials include surgeons on the learning curve, they will be assessing the learning curve as well the procedure itself. Thus clinical trials aimed at assessing the procedure should ensure that the surgeons included in the trial have all completed their learning curve. Currently, even well constructed and detailed trial protocols such as the WHiTE3 study do not address this aspect. To ensure this is the case and that it is apparent, future
surgical trial protocols should state (1) the number of cases they consider a surgeon should have performed to complete the learning curve for that procedure and (2) that all the surgeons contributing to that trial have completed that number of cases.

The second area that is known to affect outcome is case volume. Where a condition or procedure is rare then a multicentre study is entirely appropriate to ensure case ascertainment. However when a condition is common and a multicentre trial is conducted it is important that the total case burden for that condition at the local contributing site is known; to indicate that the site is experienced; and that case ascertainment is representative of the condition burden and not a very highly select subgroup thereof, particularly if the patient demographics do not match the general demographics of the study or local overall demographics for that condition. Furthermore it is important that for operative procedures, or indeed the conservative management of fractures that the expertise to perform the intervention is assessed.

Whilst considering the PROFHER study,16 Ghert and Mckee17/18 draw attention to the fact that 66 surgeons operated on 125 patients over a two and a half year period, i.e., less than one patient per surgeon per year. To judge the proficiency of the participating surgeons it would have been very helpful to have data on the number of cases they were performing in total each year. Other examples where this information would have been useful include studies looking at cemented versus uncemented hip replacement19 and at metal on metal resurfacing.19 Larger studies involving DVT prophylaxis have included recruitment in 27 countries but not specified the number of recruitment centres or the total number of the relevant arthroplasty performed at those centres.20 Thus as well as recording the usual recruitment and exclusions data it is important to understand the total population in these multicentre studies in order to draw appropriate conclusions from the data.

The potential differences in outcome of complex surgery delivered in high-volume centres as opposed to low volume centres has also been noted to be a potentially important factor in the design of pragmatic surgical trials.21

Thus future surgical trial protocols should state in the protocol (1) the minimum surgical volume considered necessary to maintain proficiency and (2) the trial data should include the number of cases surgeons are doing per year outside the study as well as those within the study.

If this data is not reported it is impossible to judge whether any inferior results are attributable to the procedure itself or to delivery of the procedure by the surgeon.

In summary, surgical trials should state; (i) in the protocol - the number of cases in the learning curve and (ii) the recommended annual surgical volume and (iii) when reporting the trial - that contributing surgeons have completed the learning curve and (iv) the number of cases of the study type performed by each surgeon both within and outside the study.

References


Funding Statement

None declared.

ICMJE Conflicts of Interest

None declared.

© 2017 Simpson, Howie and Norrie. This is an open-access article distributed under the terms of the Creative Commons Attribution licence (CC-BY-NC), which permits unrestricted use, distribution, and reproduction in any medium, but not for commercial gain, provided the original author and source are credited.