Achieving the Right Volume of Randomized Controlled Trials

Heidi Nelson, MD and Karla Ballman, PhD

Ali et al[1] in this issue of *Annals of Surgery* propose that randomized controlled trials (RCTs) are the best study design to test the efficacy of a new intervention and that RCTs should influence health care decisions and policies. We agree with this premise and applaud the authors for conducting this thoughtful study, which provides a snapshot analysis of the global status of surgical trials past and present. Findings from Ali et al are consistent with others, and so we accept the conclusions of the paper as valid. The study concludes (1) that the volume of surgical trials has increased significantly; (2) that nonindustry trials have almost doubled; and (3) that the quality of trials has improved.[2,3] We agree that RCTs are critical for the adoption of new surgical practice interventions and that monitoring progress in this domain is useful. What remains unclear is what constitutes the "right volume" of surgical RCTs; that is, as we continue to monitor the volume of surgical trials, how will we know we have achieved the optimal amount? We are not aware that anyone has considered establishing a target for the right number of surgical trials.

As a theoretical construct, the level of surgical trials should be matched by the level of surgical innovation; whether it means testing new products such as meshes, equipment such as robots, pathways such as enhanced recovery, or tools such as the electronic medical record. Even clinical practice tools that are not considered traditional clinical interventions, such as the electronic medical record, have been shown to have an impact on patient outcomes. Take for example the physician order entry system. When this tool was introduced into a pediatric intensive care unit, the unintended consequence of changes in workflow patterns resulted in a significant increase in mortality.[4] We are practicing in a time of great innovation in all aspects of surgery and medicine, and so, it could be argued that it is impractical and too costly to conduct RCTs for every new clinical tool or technique. We rebut that it is costly and unsafe to introduce new interventions without proper testing. This very point is playing out in the field of robotics where patients are suing the device company for life-threatening misadventures.[5] Given the breadth and depth of the rapidly evolving surgical innovations, there must be a logical prioritization strategy as well as diverse methods for systematically assessing new tools, techniques, and products. We offer that if 3 key factors are considered, namely magnitude of risk, prospective study design, and minimization of selection bias, then a number of different methods may be used to accurately assess new technologies and practices.

In situations where patient populations experience high rates of death or complications, proposed interventions may have a major impact, and/or proposed interventions are costly, we submit that a rigorous testing program such as used for cancer drugs should be employed. A phase I study determines the safety profile, a phase II study estimates the potential efficacy, and a phase III RCT tests the risks and benefits using standardized outcomes in defined populations. These trials incorporate monitoring mechanisms to ensure high quality, reliable data and to ensure that no undue harm comes to study participants. It is standard for these trials to use random allocation for treatment arm assignment and an intent-to-treat analytic plan. In this way, patient selection bias is eliminated and true attribution can be assigned to the superior intervention. The drawbacks of RCTs are that they may not be feasible when the 2 study interventions are markedly different (ie, local excision vs abdominal perineal resection for early-stage rectal cancer), they consume considerable time and resources to conduct, and they often do not reflect the real world clinical practice.[6]

Studies involving markedly different interventions are often not feasible because care providers are unwilling to randomize patients. An alternative study design to address this would be a cluster randomized trial, which allows a comparison between interventions where equipoise exists at an institutional level but not at the individual provider level. Although a cluster randomization design has several pitfalls,[7] it minimizes patient selection bias when properly implemented. In general, cluster randomization trials provide stronger evidence compared with registries (eg, SEER—Surveillance, Epidemiology, and End Results) for comparing patient outcomes between 2 interventions. The institutional randomization and the institutional commitment to approach all eligible patients for trial participation minimizes patient selection bias, which is inherent in patient registry databases, even those of highest quality.

From the Mayo Clinic, Department of Surgery, Rochester, MN.
Disclosure: The authors declare no conflicts of interest.
Reprints: Heidi Nelson, MD, Mayo Clinic, Department of Surgery, 200 First Street, SW, Rochester, MN 55905. E-mail: nelsonh@mayo.edu.

ISSN: 0003-4932/13/25802-0208
DOI: 10.1097/SLA.0b013e31829c4a05

As pointed out in the article by Ali et al, the RCT is costly for a number of reasons including the expense of data collection. We tested the possibility of using existing Commission on Cancer (COC) Cancer Registries to obtain long-term cancer outcomes of survival and recurrence for a clinical trial. Within our own institution, data from our COC Cancer Registry provided the same results as the data collected by 2 different National Cancer Institute Cooperative Group trials.[8] If this observation could be replicated by other COC centers, it would be possible to rely on Cancer Registries to obtain long-term patient outcomes and substantially reduce the cost of an RCT. For short-term endpoints such as postoperative complications, the utilization of well-established national databases (eg, American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP®)) might reduce the burden of data collection for institutions that participate in these initiatives.

For testing low-risk, low-cost incremental advances, it may be sufficient to utilize prospective practice-based registries or smaller randomized trials to obtain necessary evidence. Suppose some evidence exists that a low-risk, low-cost intervention has favorable patient outcomes. A possible confirmatory study design would be to implement the intervention across the practice and then compare patient outcomes immediately before the change in practice with those immediately after the change, a before-and-after design. The requirements for a high-quality before-and-after design are (1) a prospective plan with a predetermined sample size to assure adequate precision for the impact on patient outcomes, (2) a globally introduced change to the practice, to eliminate patient selection bias, and (3) a standardized, audited database to capture outcomes of interest. An added benefit is that a before-and-after design provides better estimates of "real world" impact than an RCT. An alternative approach is a smaller, less rigorous RCT, which would provide an imprecise estimate of the potential impact of an intervention due to its small sample size. The benefits of smaller RCTs are that they are less costly, provide good quality information, and can be combined with other small RCTs via a meta-analysis to gain a more precise estimate of potential impact. The small RCT can also provide early evidence of futility, that is, evidence of the intervention does not have the anticipated impact on patient outcomes, thus saving the expense of a large RCT. The major drawback of a small RCT is that it is underpowered; if a small RCT does not yield a statistically significant difference, it could be that a clinically meaningful difference has been missed.

In closing, we know from this month's report in *Annals of Surgery* that we are all moving in the right direction, but that Europe has made greater progress in increasing the relative volume of RCTs, which they credit to their mature research infrastructure. In the United States, we do not have a single mature research infrastructure, but we do have professional organizations, such as the American College of Surgeons, investing in national surgical databases, and surgical societies investing in the development of practice guidelines and the formation of procedure-specific study groups. In that regard, we are well positioned to take advantage of many resources and study design methods as outlined earlier to set and achieve real volume targets for testing surgical innovations. What remains is for the study groups to find sources of funding that can provide sustainable infrastructure. Achieving sustainable funding will be the greatest challenge and a topic deserving of its own editorial.

REFERENCES

1. Ahmed Ali U, van der Sluis PC, Issa Y, et al. Trends in worldwide volume and methodological quality of surgical randomized controlled trials. *Ann Surg*. 2013;258:199–207.
2. Hopewell S, Dutton S, Yu LM, et al. The quality of reports of randomized trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ*. 2010;340:c723.
3. Brooke BS, Nathan H, Pawlik TM. Trends in the quality of highly cited surgical research over the past 20 years. *Ann Surg*. 2009;249:162–167.
4. Han YY, Carcillo JA, Venkatraman ST, et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics*. 2005;116:1506–1512.
5. Rabin RC. Salesmen in the surgical suite. *New York Times*. March 25, 2013:D1.
6. Chow CJ, Habermann EB, Abraham A, et al. Does enrollment in cancer trials improve survival? *J Am Coll Surg*. 2013;216:774–780.
7. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Public Health*. 2004;94:416–422.
8. Shi Q, You YN, Nelson H, et al. Cancer registries: a novel alternative to long-term clinical trial follow-up based on results of a comparative study. *Clin Trials*. 2010;7:686–695.