Well-designed, well-executed, and well-reported randomized clinical trials (RCTs) are the cornerstone of improved care in medicine. While RCTs sit atop the hierarchy of evidence in resolving questions of therapy, they are predicated on the critical assumption of what is known as external validity, i.e., are the study methods of sufficiently high quality to warrant their application to the routine care of patients? The potential for various forms of bias in the trial process is an Achilles heel threatening trial validity and interfering with the optimal uptake of negative or positive trial findings.

Trials of surgical interventions are especially challenged by the ostensible inability to conduct a trial with double-blind randomization. Robinson et al, in their article “Characteristics of Randomized Clinical Trials in Surgery From 2008 to 2020: A Systematic Review,” use various tools and approaches to summarize key quality measures for all surgical trials (n = 388) published in 8 leading journals during the study period. We highlight a sample of these. Using the Cochrane Risk of Bias 2 Tool (CRB2T), they demonstrated that 211 trials (54.4%) and 91 trials (23.5%) had a moderate and high risk of bias, respectively. The CRB2T has 5 demarcated domains, including randomization process, deviations from intended interventions, missing outcomes data, measurement of outcomes, and selection of reported results. Overall, 85 trials (21.9%) had a high risk of bias for the selection of reported results. Furthermore, most trials (303 [78.1%]) did not control for surgeon experience, and only 17 trials (4.4%) assessed the quality of the intervention.

There are solutions for issues highlighted in this review; some can be easily implemented, while others likely require surgical culture shifts. Preregistration of trials, now required by most journals, should minimize many biases, especially selection of reported results. Toward the last years of publication used in the review by Robinson et al, nearly 90% of trials were preregistered, suggesting that today the diligent reviewer or reader guided by a tool such as the CRB2T can quickly consider trial biases and act accordingly. An expertise-based trial design can address the issue of surgical learning curves and intervention fidelity for innovation trials. Currently, innovation trials are usually led by innovation enthusiasts who provide both the innovation and control procedures (e.g., laparoscopic and open surgery) and potentially influence patient outcomes (e.g., decisions for discharge date). A negative trial should not change practice. However, a positive innovation trial may be open to charges of suboptimal quality of care in the control group and lack of blinding bias. In an expertise-based trial, patients are enrolled and only then randomized to expert group A (surgeons proficient in the innovation) or expert group B (surgeons proficient in the control procedure). The use of separate surgical teams masked to intervention group and providing postoperative care should mitigate lack of blinding bias. Patient blinding can be challenging but achieved in appropriate circumstances through extensive dressings or even sham surgery trial designs.

Current models of care largely preclude a surgical expertise trial design. It is our observation that an initial consultation with a surgeon often creates some form of imprinting for a patient, making them reluctant to be treated by a different surgeon for the presenting concern. An expertise-based trial requires marked cooperation and coordination among surgeons who may have substantial differences in equipoise toward the intervention under review. The willingness of surgeons to participate in an expertise-based trial would represent a cultural shift, would reflect true patient-centered care, and could be of use in many areas of surgery. These last comments are likely also relevant to the observation that most trials in the review originated in Europe, with fewer than...
one-fifth in North America. This suggests that in these 2 jurisdictions, differences in surgical culture result in a varying overall commitment to the execution of surgical RCTs.

This review encourages comments on trial reproducibility. In psychology, leaders call the inability to reproduce findings of many historic sentinel trials a reproducibility crisis. In surgical RCTs, the more important issue may be a lack of reproducing previous sentinel trials using contemporary diagnostic, therapeutic, and methodologic standards.4 For example, a trial by Patchell et al.5 published in 1990, randomized 48 patients with a single brain metastasis to surgery and postoperative whole brain radiation (n = 25) vs biopsy and whole brain radiation (n = 23).5 Findings of this single small RCT still largely inform treatment recommendations for patients with accessible single brain metastases to undergo surgery, despite numerous advances since the study was conducted (eg, brain magnetic resonance imaging, gamma knife radiation, and chemotherapeutic and biologic therapeutic innovations). It would be ideal to replicate this trial with updated methods and better parse out which intersection of patient and tumor characteristics should result in specific treatment recommendations.

Only findings from rigorous and accurate surgical RCTs should be added to the totality of evidence in an area and influence subsequent patient care. The goal of ensuring trial reproducibility and external validity is paramount. The thorough review by Robinson et al1 found many concerns regarding surgical RCTs and should serve as a challenge to all stakeholders interested in improving surgical care. Shifts in surgical culture are likely necessary to ensure it is the rule and not the exception that a patient undergoing surgery is enrolled in a well-designed and well-executed RCT.

REFERENCES