The two Achilles heels of surgical randomized controlled trials: differences in surgical skills and reporting of average performance

Farr R. Nezhat, MD; Cande V. Ananth, PhD, MPH; Anthony M. Vintzileos, MD

Randomized controlled trials of surgery are fundamentally different from randomized controlled trials of medications because it is difficult to blind or mask a surgical procedure or perform "sham" operations. An additional challenge is the variation in skills and surgical proficiency of participating centers and surgeons. Addressing heterogeneity in surgical proficiency remains of paramount importance, especially when randomized controlled trials involve a new or complex procedure such as minimally invasive radical surgery. In the presence of such heterogeneity, it is very cumbersome to evaluate objectively and monitor surgical skills so that most trials simply report associations that are averaged across surgeons and hospitals/centers. Such reporting is not transparent because the rates of complications and adverse outcomes are reported only as averages, and these averages may not apply to the individual participating surgeons or centers. These factors, coupled with the inherent nongeneralizability of findings from such randomized controlled trials, because of the strict inclusion and exclusion criteria for enrollment, may lead to conclusions that no longer apply to real life for individual surgeons or centers. Case in point is a recently published noninferiority randomized controlled trial that reported that minimally invasive radical hysterectomy was associated with lower rates of disease-free survival (86% vs 96.5% at 4.5 years) and overall survival (93.8% vs 99% at 3 years) than open abdominal radical hysterectomy in patients with cervical cancer. However, randomized controlled trials that involve 2 competing complex or new procedures may be affected by tremendous confounding because of variations in surgical proficiency and also nonstandardization for other confounding factors such as patient selection categories (ie, stage of cancer) and adjuvant postoperative therapies that may affect long-term survival. The purpose of this Viewpoint is not to provide an exhaustive review of the trial's shortcomings but to use it as an illustration to focus on 2 challenging areas that most randomized controlled trials of a new complex surgical procedure suffer from: (1) unadjusting or not correcting for surgical skill variability and (2) nontransparent reporting of averaged results. We provide suggestions to overcome these deficiencies through robust methods and statistical approaches.

Key words: cervical cancer, laparotomy, nontransparent reporting, randomized trials, surgical learning curve, surgical proficiency

The central dogma of all research is to benefit patients first and foremost. The challenges in the design, conduct, and analysis of surgical randomized controlled trials (RCTs) have been described previously in detail.[1] We reported the first laparoscopic radical hysterectomy with paraaortic and pelvic node dissection in 1992.[2] Since then, several groups have reported equivalent disease-free survival and overall survival for this procedure compared with laparotomy in addition to benefits such as shorter hospital length of stay, less blood loss, and lower short- and long-term morbidity.[3–14] In contrast, a recently reported RCT by Ramirez et al[15] concluded that there was a lower rate of disease-free survival and overall survival with minimally invasive radical hysterectomy compared with open abdominal radical hysterectomy among women with early-stage cervical cancer. However, RCTs that involve 2 competing complex or new procedures may be affected by confounding bias because of variations in surgical skills of both the participating surgeons and the surgical proficiency of the entire surgical team at each center.

Adjusting for surgical skills

The conclusion of the aforementioned RCT that involved an unknown number of surgeons that was conducted in 33 centers across 13 countries and 5 continents over a 9-year period should be viewed with skepticism. Three hundred nineteen patients underwent minimally invasive radical hysterectomy. If we assume that 2–3 surgeons participated from each center, then the average contribution of study cases is estimated at 3–5 cases per surgeon over the 9 years or approximately 1 study case (1 minimally invasive radical hysterectomy) per center per year. Previous research has shown that low-volume surgery is associated with less technical skill and higher complication rates.[16,17]

From the Department of Obstetrics and Gynecology, New York University (NYU) Winthrop University Hospital, Mineola, NY (Drs Nezhat and Vintzileos); the Department of Obstetrics and Gynecology, Weill Cornell Medical College of Cornell University, NY (Dr Nezhat); the Division of Epidemiology and Biostatistics, Department of Obstetrics, Gynecology, and Reproductive Sciences, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ (Dr Ananth); the Environmental and Occupational Health Sciences Institute, Rutgers Robert Wood Johnson Medical School, Piscataway, NJ (Dr Ananth).

The authors report no conflict of interest.

Corresponding author: Farr R. Nezhat, MD. farr@farrnezhatmd.com

0002-9378/$36.00 • © 2019 Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.ajog.2019.05.017

None of the 33 participating centers or surgeons had to exhibit surgical proficiency in performing the traditional procedure (ie, open radical hysterectomy) other than a track record in clinical research and hospital privileges. However, these qualifications do not guarantee surgical proficiency or equivalency in surgical skills among the participating surgeons. The investigators tested for surgical proficiency only for minimally invasive radical hysterectomy by reviewing the outcomes of at least 10 cases and 1 video per surgeon at the beginning of the trial,[18] with the protocol changing to 10 cases and 2 videos per center during the trial.[15] The trial management committee did not rank the centers or surgeons according to their submitted outcomes, nor did they use an objective evaluation or scoring of the main domains of their technical skills on tape review (eg, gentleness, tissue exposure, instrument handling, time and motion, and flow of operation).[16] As a matter of fact, the protocol did not call for any standardization, documentation, or monitoring of the necessary surgical steps to ensure that the procedure was accomplished appropriately as is recommended.[19] In addition, clustering by center or surgeon and need for randomization stratified by center (or surgeon) were not considered.[20] In our view, the surgical skill reporting and standardization performed were insufficient to ensure or to assess surgical proficiency, which may have led to average reporting of results as the only option. It should be noted that no data were provided regarding the number and reasons for excluding centers or surgeons or if any exclusions occurred because of suboptimal surgical outcomes or technical skills after case and video review.

Before the initiation of the trial, the expectation of the investigators was that most patients would come from US institutions.[18] However, only 6 of 33 participating centers were from the United States at the end of the trial. The investigators did not provide an explanation for this discrepancy, nor did they disclose whether there were difficulties in recruiting patients initially that required them to relax the initial (2008) technical skills criteria protocol[18] or to open the trial to more centers outside the United States.

One of the generally accepted advantages of minimally invasive surgery, as compared with open laparotomy, is the expected shorter hospital stay, improved patient quality of life, lower treatment-related morbidity, and lower early postoperative complication rate.[3−14,21] In the early cervical cancer RCT, however, early postoperative complication rates were similar between minimally invasive and open radical hysterectomy (25% vs 26%, respectively)[15] that cast doubt about the generalizability of other more primary outcomes; additionally, the finding that locoregional recurrences were more frequent in the minimally invasive group raises concern that suboptimal technique may be responsible for recurrences rather than the route of the surgical procedure.

Nontransparent reporting of the results

Most multicenter surgical trials report outcomes and complications that are averaged across centers without consideration of rates of individual centers or surgeons. Although this may be considered acceptable for so-called pragmatic trials,[1] such reporting is not very helpful for surgical trials if expertise or volume varies widely by center or surgeon. More meaningful reporting in such cases is possible after an appropriate skill evaluation and stratification of results according to individual centers by size, volume, and/or surgical skills. In minimally invasive surgery, it has been shown that surgical outcomes vary with technical skills; surgeons at the lowest technical skill score quartile have higher mortality and complication rates, longer operations and higher rates of reoperation and readmission, as compared with those at the upper quartile.[16]

The early-stage cervical cancer trial lasted 9 years (2008−2017); in trials with a long duration, results should be compared in the most recent years to the initial years of the study because of the surgical learning curve.[1,20] No data on the learning curve were reported; instead, the authors reported that cervical cancer recurrences occurred in 14 of 33 recruiting centers but provided no data regarding the individual centers that had recurrence-free outcomes and whether these outcomes were associated with center or surgeon characteristics, such as case volume.

Unfortunately, aggregated reporting remains entrenched in surgical intervention trials. For instance, another multicenter RCT that compared laparoscopy with laparotomy for uterine cancer staging reported a laparotomy conversion rate of 14.6% among those who were assigned randomly to laparoscopy (246/1682) because of "poor visibility."[22] However, the laparotomy conversion rates, when analyzed by institution size, ranged from as low as 14.9% to as high as 34.7%, despite similarities in the patients' body mass indices.[22] This trial did not attempt to adjust for or evaluate the surgical proficiency of the participating surgeons or centers. Therefore, the heterogeneity in laparotomy conversion rates may reflect large variation in surgical proficiency among the participating institutions.

The solution

The lack of transparency in the reporting of the results of surgical RCTs has its origin in inappropriate method and especially lack of meaningful evaluation and adjustment for surgical proficiency. Given the lack of evaluation of surgical proficiency, we believe that outcome reporting by individual centers or surgeons may be more informative. Currently, hospitals and healthcare organizations are asked routinely and do report their outcomes to the public and governmental agencies for comparison with other hospitals in patient care quality metrics and for reimbursement reasons. There is no reason that we cannot do the same in our reporting of surgical RCTs.

To address the heterogeneity in surgical outcomes across centers and surgeons, we propose a multilevel modeling framework to estimate rates by center (surgeon).[23] One modeling approach is a random-intercept generalized linear model. In this model, each center is introduced as a random intercept, which enables outcomes to vary across centers. A slightly expanded random-intercept model is 1 in which surgeons can be "nested" within a center so that outcomes are allowed to vary not only across centers but also by surgeons within centers. This model will address explicitly the variability phenomenon and report ranges of surgical outcomes that are displayed by center and/or surgeon.

Comment

Future surgical RCTs should be designed carefully to determine, monitor and document surgical skills objectively, to consider learning curve and stratified randomization, and to report the variation in results according to surgical skills or volume. Transparency of the results should be improved by reporting not only averages but also the specific outcomes of each participating center after appropriate statistical modeling. Reporting by quartile skill, center size, or individual center reporting may lead to reduced power (requiring larger samples) but at least will offer readers and patients more meaningful results. Averaged reporting without considering the variations in surgical proficiency could result in unforeseen adverse consequences, such as the recently released Food and Drug Administration statement that used the trial by Ramirez et al[15] as evidence to caution against robot-assisted surgical devices for treatment of patients with cancer.[24] It is our hope that, contrary to the Food and Drug Administration cautionary statement, minimally invasive surgical approach should not be replaced by open surgery based on the averaged results of the Ramirez et al trial. ■

REFERENCES

1. Cook JA. The challenges faced in the design, conduct and analysis of surgical randomized controlled trials. Trials 2009;10:9.
2. Nezhat CR, Burrell MO, Nezhat FR, Benigno BB, Welander CE. Laparoscopic radical hysterectomy with paraaortic and pelvic node dissection. Am J Obstet Gynecol 1992;166:864–5.
3. Zakashansky K, Chuang L, Gretz H, Nagarsheth NP, Rahaman J, Nezhat FR. A case-controlled study of total laparoscopic radical hysterectomy with pelvic lymphadenectomy versus radical abdominal hysterectomy in a fellowship training program. Int J Gynecol Cancer 2007;17:1075–82.
4. Lee EJ, Kang H, Kim DH. A comparative study of laparoscopic radical hysterectomy with radical abdominal hysterectomy for early-stage cervical cancer: a long-term follow-up study. Eur J Obstet Gynecol Reprod Biol 2011;156:83–6.
5. Nam JH, Park JY, Kim DY, Kim JH, Kim YM, Kim YT. Laparoscopic versus open radical hysterectomy in early-stage cervical cancer: long-term survival outcomes in a matched cohort study. Ann Oncol 2012;23:903–11.
6. Steed H, Rosen B, Murphy J, Laframboise S, De Petrillo D, Covens A. A comparison of laparoscopic-assisted radical vaginal hysterectomy and radical abdominal hysterectomy in the treatment of cervical cancer. Gynecol Oncol 2004;93:588–93.
7. Frumovitz M, dos Reis R, Sun CC, et al. Comparison of total laparoscopic and abdominal radical hysterectomy for patients with early-stage cervical cancer. Obstet Gynecol 2007;110:96–102.
8. Malzoni M, Tinelli R, Cosentino F, Fusco A, Malzoni C. Total laparoscopic radical hysterectomy versus abdominal radical hysterectomy with lymphadenectomy in patients with early cervical cancer: our experience. Ann Surg Oncol 2009;16:1316–23.
9. Soliman PT, Frumovitz M, Sun CC, et al. Radical hysterectomy: a comparison of surgical approaches after adoption of robotic surgery in gynecologic oncology. Gynecol Oncol 2011;123:333–6.
10. Sert BM, Boggess JF, Ahmad S, et al. Robot-assisted versus open radical hysterectomy: a multi-institutional experience for early-stage cervical cancer. Eur J Surg Oncol 2016;42:513–22.
11. Shah CA, Beck T, Liao JB, Giannakopoulos NV, Veljovich D, Paley P. Surgical and oncologic outcomes after robotic radical hysterectomy as compared to open radical hysterectomy in the treatment of early cervical cancer. J Gynecol Oncol 2017;28:e82.
12. Shazly SA, Murad MH, Dowdy SC, Gostout BS, Famuyide AO. Robotic radical hysterectomy in early stage cervical cancer: a systematic review and meta-analysis. Gynecol Oncol 2015;138:457–71.
13. Wang YZ, Deng L, Xu HC, Zhang Y, Liang ZQ. Laparoscopy versus laparotomy for the management of early stage cervical cancer. BMC Cancer 2015;15:928.
14. Cao T, Feng Y, Huang Q, Wan T, Liu J. Prognostic and safety roles in laparoscopic versus abdominal radical hysterectomy in cervical cancer: a meta-analysis. J Laparoendosc Adv Surg Tech A 2015;25:990–8.
15. Ramirez PT, Frumovitz M, Pareja R, et al. Minimally invasive versus abdominal radical hysterectomy for cervical cancer. N Engl J Med 2018;379:1895–904.
16. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med 2013;369:1434–42.
17. Mowat A, Maher C, Ballard E. Surgical outcomes for low-volume vs high-volume surgeons in gynecology surgery: a systematic review and meta-analysis. Am J Obstet Gynecol 2016;215:21–33.
18. Obermair A, Gebski V, Frumovitz M, et al. A phase III randomized clinical trial comparing laparoscopic or robotic radical hysterectomy with abdominal radical hysterectomy in patients with early stage cervical cancer. J Minim Invasive Gynecol 2008;15:584–8.
19. Blencowe NS, Mills N, Cook JA, et al. Standardizing and monitoring the delivery of surgical interventions in randomized clinical trials. Br J Surg 2016;103:1377–84.
20. Conroy EJ, Rosala-Hallas A, Blazeby JM, Burnside G, Cook JA, Gamble C. Randomized trials involving surgery did not routinely report considerations of learning and clustering effects. J Clin Epidemiol 2019;107:27–35.
21. Kornblith AB, Huang HQ, Walker JL, Spirtos NM, Rotmensch J, Cella D. Quality of life of patients with endometrial cancer undergoing laparoscopic international federation of gynecology and obstetrics staging compared with laparotomy: a gynecologic oncology group study. J Clin Oncol 2009;27:5337–42.
22. Walker JL, Piedmonte MR, Spirtos NM, et al. Laparoscopy compared with laparotomy for comprehensive survival staging of uterine cancer: gynecologic oncology group study LAP2. J Clin Oncol 2009;27:5331–6.
23. McCullagh P, Nelder JA. Generalized linear models. New York: Chapman & Hall/CRC Monographs on Statistics and Applied Probability; 1983.
24. Caution when using robotically-assisted surgical devices in women's health including mastectomy and other cancer-related surgeries: FDA safety communication February 28, 2019. Available at: https://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/ucm632142.htm. Accessed March 4, 2019.