

Chapter 28

Surgeon Training and the Learning Curve

Kamal M.F. Itani

Introduction

With every intervention, procedure, interpretation of a test or even new medical treatment, there is a learning curve. This learning curve differs among providers based on background, training, skills, environment, and available support as well as the similarity of the new intervention to older ones. In addition, moving beyond the learning curve to experienced provider in a new intervention is open to interpretation based on the observer, his/her background within the field, and the observer status on the learning curve for that intervention.

When testing a new intervention in a prospective randomized trial, the investigators will have to decide on the level of expertise of each participating investigator, their standing on the learning curve, the level of expertise that each has to achieve prior to enrolling patients, and how it is measured and ethical considerations related to patients and society. This chapter will discuss each of these points and how to address them within a large prospective randomized clinical trial.

Definition

The learning curve has been defined as the time it takes and/or the number of procedures an average surgeon needs to perform independently in order to reach a reasonable outcome [1]. Others have represented the learning curve as the relationship between experience with a new procedure or technique and an outcome

K.M.F. Itani (✉)

Department of Surgery, VA Boston Health Care System, Boston University
and Harvard Medical School, VABHCS(112A), 1400 VFW Parkway,
West Roxbury, MA 02132, USA
e-mail: kitani@va.gov

© Springer International Publishing AG 2017

K.M.F. Itani and D.J. Reda (eds.), *Clinical Trials Design in Operative
and Non Operative Invasive Procedures*, DOI 10.1007/978-3-319-53877-8_28

variable such as operating time, complication rate, hospital stay, or mortality [2]. A learning curve may also be operationally defined as an improvement in performance over time. It therefore implies a baseline performance, an improvement over time which can happen at various rates of speed, and a plateau in performance afterwards. The speed with which a plateau is achieved is dependent on the initial performance level and the rapidity with which the improvement occurs up to the plateau. It is to note that depending on the learning curve phase, lack of investigator equipoise might exist favoring traditional interventions during the baseline or improvement phase and possibly favoring the newer intervention during the plateau phase.

(a) Baseline performance

Baseline performance depends on the individual baseline skills and familiarity with similar interventions or exposure to similar interventions in the past. For example, an orthopedic surgeon performing hip or knee replacements might be comfortable with one or two prostheses that are commonly used. However, when a new prosthesis is introduced into practice, it might require a new set of skills some of which overlap with the old ones and some which are totally new. The level of overlap is also dependent on the type of prosthesis the surgeon was using.

(b) The improvement phase

The improvement phase is also dependent on each individual surgeon's background with the technology, learning abilities, as well as the environment in which they practice. The environment might have other experts able to provide feedback about progress, a larger volume of patients to be treated with the newer intervention, the availability of cadavers, animal labs, or simulators to practice. All of these will factor into the speed at which the plateau is reached.

(c) Plateau phase

During the plateau phase, the individual is considered familiar, comfortable, and experienced in performing the newer intervention and should be able to teach it to others interested in acquiring these skills. The assessment that the individual is at the plateau phase is arbitrary and can be a function of reported volume, time, observation, or a combination of all the above. Any auditor of this new technology should be at the plateau phase.

Selection of Investigators in a Trial

Participation in clinical trials requires that investigators have proven capability and knowledge in the conduct of the research-related operations and ideally at the plateau phase. Parameters for how many operations the surgeon is required to perform must be established with assessment of a defined outcome measure. In

some cases, certification might be done by submitting a record of operative and pathology results. Alternatively, the surgeon might be required to submit videos that could be audited and reviewed. For other operations, observation by a proctor can confirm that the surgeon is ready to perform the operation as part of a clinical trial. Techniques and operations that are already part of the surgeon's skills still need to be assessed to measure the surgeon's ability to perform the operation in accordance with the requirements of the study. As mentioned in the chapters on investigators meetings (Chap. 37) and site visits (Chap. 38), the principal investigator must budget for training, including providing funds to train sub-investigators, stipends, and travel for proctors.

In the prospective randomized trial comparing laparoscopic to open inguinal hernia with mesh, the laparoscopic technique was relatively new and had little to moderate penetration in clinical practice. It was arbitrarily decided with the help of experts that each surgeon participating in the trial should have a minimum experience of 25 cases performed laparoscopically, a videotaped laparoscopic repair reviewed by the principal investigator before the start of the study, and random videos of the procedure sent for auditing during the course of the study [3]. A post hoc analysis of the data looking at the influence of volume, age, and time since board certification revealed that a volume of 250 cases was necessary to achieve with the laparoscopic repair the same level of recurrence and complications as the open repair [4]. It became clear that most surgeons participating in the trial were still in the improvement phase. This trial clearly demonstrated the steep learning curve for performing laparoscopic inguinal hernia repairs and that these operations are best performed by surgeons at the plateau phase of the learning curve in order to achieve the desired outcome.

In the prospective randomized trial evaluating laparoscopic assisted vs. open resection of stage 2 or 3 rectal cancer on pathologic outcomes [5], a credentialing committee was established to review unedited videotapes, operative reports, and pathology reports of 46 participating surgeons at 35 institutions. The mechanism of credentialing of participating surgeons is described in an online supplement of the published study [6]. In addition a random audit performed for the first 100 laparoscopic cases was confirmatory of expertise in techniques used throughout the trial. The principal investigators wanted to ensure that all participating surgeons in this trial were at least at the plateau phase for each of the surgeries included in that trial. This resulted in a 93% compliance in the rate of total mesorectal excision, an important goal of the study and a reflection of the high quality of surgery performed in that study.

Auditing Results

It is the responsibility of the principal investigator and the executive committee to audit and monitor any new surgical procedure. Stopping rules must be in place to remedy or remove a surgeon who is not performing as expected with regard to the

technical requirements of the intervention or if placing patients at risk. A priori decisions need to be made with regard to whether the data from surgeons who are removed will be used in the final analysis of the results.

Auditing can occur through live visits to investigator sites, videotapes, or close monitoring of outcomes.

Statistical Considerations

Various statistical methods have been reported in the assessment of the learning curve [7]. Commonly used data are split into arbitrary groups and the means compared by chi-squared test or ANOVA. Some studies have data displayed graphically with no statistical analysis. Others use univariate analysis of experience versus outcome. Some studies use multivariate analysis techniques such as logistic regression and multiple regression to adjust for confounding factors. A systematic review concluded that the statistical methods used for assessing learning curves have been crude and the reporting of studies poor [8]. Recognizing that better methods may be developed in other nonclinical fields where learning curves are present (psychology and manufacturing), a systematic search was made of the nonclinical literature to identify novel statistical methods for modeling learning curves. A number of techniques were identified including generalized estimating equations and multilevel, or hierarchical, models. The main recommendation was that given the hierarchical nature of the learning curve data and the need to adjust for covariates, hierarchical statistical models should be used. Ramsay et al. [8] went further to suggest Bayesian hierarchical modeling in order to adjust for effect sizes for learning.

Biau et al. [9] suggested the cumulative summation test for learning curve that allows quantitative and individual assessment for the learning curve. The cumulative summation test has been applied to the learning curve and is designed to indicate when a process deviates from an accepted level of performance.

Other statistical tools available to address the learning curve include the intraclass correlation coefficient. In multicenter trials, data from the same center are more similar than those from different centers. These similarities which often include the level of the center as a whole on the learning curve induce a correlation between data, known as the center effect. This center effect is assessed by the intraclass correlation coefficient [10].

Finally, an expertise-based approach to trial design, where health professionals only deliver an intervention in which they have expertise, has been proposed as an alternative. An expertise-based trial design should be considered but its value seems context-specific, particularly when the control and the intervention under study differ substantially or are delivered by different health professionals [11].

Ethical Consideration

There is no doubt that patients undergoing procedures earlier during the learning curve are at greater risk for adverse events than patients operated on during the plateau phase of the learning curve or by experienced surgeons. Information related to the new procedure and the learning curve of the surgeon should be included in the informed consent and discussed with the patient. This can by itself bias the patient against the procedure and preclude a subject from participating in a trial for fear of undergoing a new procedure by a less-experienced surgeon.

There is also the dilemma of potentially promising procedures that are difficult to learn, replicate, or teach. Such procedures are of limited generalizability and risky to test within the context of a multicenter trial.

Conclusion

When testing a new intervention, principal investigators have a responsibility to evaluate the learning curve and the status of each investigator on the learning curve. Participating subjects should be informed of the expertise level of the investigator caring for them. Lack of equipoise among investigators or bias against a new procedure by patients may be introduced based on the learning curve of the investigator. Frequent monitoring and auditing should be in place to avoid exposing patients to risk and compromise the results of the trial. Various statistical tools are also available to address variability resulting from the learning curve.

References

1. Subramonian K, Muir G. The learning curve in surgery: what is it, how do we measure it and can we influence it? *BJU Int.* 2004;93(9):1173–4.
2. Michel LA. Epistemology of evidence-based medicine. *Surg Endosc.* 2007;21(2):145–51.
3. Neumayer L, Giobbie-Hurder A, Jonasson O, Fitzgibbons R Jr, Dunlop D, Gibbs J, Reda D, Henderson W. Veterans affairs cooperative studies program 456 investigators: open mesh versus laparoscopic mesh repair of inguinal hernia. *N Engl J Med.* 2004; 29;350(18):1819–27.
4. Neumayer LA, Gawande AA, Wang J, Giobbie-Hurder A, Itani KM, Fitzgibbons RJ Jr, Reda D, Jonasson O. CSP #456 investigators: proficiency of surgeons in inguinal hernia repair: effect of experience and age. *Ann Surg.* 2005;242(3):344–8.
5. Fleshman J, Branda M, Sargent DJ, Boller AM, George V, Abbas M, Peters WR Jr, Maun D, Chang G, Herline A, Fichera A, Mutch M, Wexner S, Whiteford M, Marks J, Birnbaum E, Margolin D, Larson D, Marcello P, Posner M, Read T, Monson J, Wren SM, Pisters PW, Nelson H. Effect of laparoscopic-assisted resection vs open resection of stage II or III rectal cancer on pathologic outcomes: the ACOSOG Z6051 randomized clinical trial. *JAMA.* 2015;314(13):1346–55.

6. Fleshman J, Branda M, Sargent DJ, et al. Effect of laparoscopic-assisted resection vs open resection of stage II or III rectal cancer on pathologic outcomes: the ACOSOG Z6051 randomized clinical trial. *JAMA*. 2015 Oct 6;314(13): e-appendix B.
7. Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Statistical assessment of the learning curves of health technologies. *Health Technol Assess*. 2001;5: 1–79.
8. Cook JA, Ramsay CK, Fayers P. Statistical evaluation of learning curve effects in clinical trials. *Clin Trials*. 2004;1:421–7.
9. Biau DJ, Porcher R, Boutron I. The account for provider and center effects in multicenter interventional and surgical randomized controlled trials is in need of improvement: a review. *J Clin Epidemiol*. 2008;61(5):435–9.
10. Vierron E, Giraudeau B. Sample size calculation for multicenter randomized trial: taking the center effect into account. *Contemp Clin Trials*. 2007;28(4):451–8.
11. Cook JA, Elders A, Boachie C, Bassinga T, Fraser C, Altman DG, Boutron I, Ramsay CR, MacLennan GS. A systematic review of the use of an expertise-based randomised controlled trial design. *Trials*. 2015;16(241):1–10.