Assessment of the learning curve in health technologies: a systematic review

Journal Item

International Journal of Technology Assessment in Health Care, **16:4** (2000), 1095–1108.

ASSESSMENT OF THE LEARNING CURVE IN HEALTH TECHNOLOGIES

A Systematic Review

Craig R. Ramsay
Adrian M. Grant
Sheila A. Wallace
Paul H. Garthwaite
University of Aberdeen

Andrew F. Monk
Ian T. Russell
University of York

Abstract

Objective: We reviewed and appraised the methods by which the issue of the learning curve has been addressed during health technology assessment in the past.
Method: We performed a systematic review of papers in clinical databases (BIOSIS, CINAHL, Cochrane Library, EMBASE, HealthSTAR, MEDLINE, Science Citation Index, and Social Science Citation Index) using the search term "learning curve."
Results: The clinical search retrieved 4,571 abstracts for assessment, of which 559 (12%) published articles were eligible for review. Of these, 272 were judged to have formally assessed a learning curve. The procedures assessed were minimal access (51%), other surgical (41%), and diagnostic (8%). The majority of the studies were case series (95%). Some 47% of studies addressed only individual operator performance and 52% addressed institutional performance. The data were collected prospectively in 40%, retrospectively in 26%, and the method was unclear for 31%. The statistical methods used were simple graphs (44%), splitting the data chronologically and performing a *t* test or chi-squared test (60%), curve fitting (12%), and other model fitting (5%).
Conclusions: Learning curves are rarely considered formally in health technology assessment. Where they are, the reporting of the studies and the statistical methods used are weak. As a minimum, reporting of learning should include the number and experience of the operators and a detailed description of data collection. Improved statistical methods would enhance the assessment of health technologies that require learning.

Keywords: Learning, Clinical competence, Technology assessment, Biomedical models, Statistical

Many innovators and early enthusiasts are reluctant to apply rigorous evaluation to a new health technology as it is being introduced into clinical practice. They argue that assessment

This project was funded by the NHS R&D Health Technology Assessment Programme. The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Executive Health Department. The views expressed are those of the authors.

at this time gives a distorted picture that is biased against the new technology (8). By the time the technology has stabilized, however, these same people are often convinced of the worth of the technology on the basis of poor evidence. Then they argue that rigorous evaluation is unethical if it involves withholding the technique or procedure from potential patients (40). This paradox has become known as Buxton's Law:

> It is always too early [for rigorous evaluation] until, unfortunately, it's suddenly too late. (11)

The basic problem is the tendency for the performance of many nondrug technologies to change over time, especially initially (46). Further technical development may continue after clinical introduction, operators may improve as a technique becomes more familiar, and performance may be enhanced by infrastructure changes, such as better trained assistants or better organized facilities (8;13). These changes generally lead to improvements in performance, and the term *learning curve* is commonly used to describe this phenomenon.

The problem is likely to continue until reliable ways are available for describing the learning curve. The British Medical Research Council identified the need for formal statistical methods in 1994 (37), but there has been little progress. This is why we systematically reviewed and appraised the methods by which the learning curve has been addressed during health technology assessment in the past.

METHODS

Search Strategy for the Identification of Studies

The search strategy was first developed in MEDLINE by a statistician and a researcher experienced in literature searching. Search terms were developed from Medical Subject Headings (MeSH) terms, using the MeSH tree with scope notes and permuted index. We also employed textword searching, that is searching for terms in the title and abstract, using truncation and adjacency where appropriate.

The number of abstracts retrieved for each term was noted, and the first 50 were assessed for relevance to learning curves. To optimize the return on resources available, a focused search strategy was then developed. The most specific search term was chosen for use in MEDLINE (Appendix 1). Other less-specific terms were tested but rejected because too many irrelevant studies were retrieved (Appendix 1). The search strategy imposed no language or similar limitations.

The search strategy was modified for other databases. The syntax was changed to suit that of the relevant search software, and the thesaurus or indices of each database were used to identify equivalents of the MeSH terms used in MEDLINE (Appendix 1). Other terms were also tested in each of the other databases but were found to be less specific and thus rejected (Appendix 1). Search terms describing complex statistical techniques that may have been appropriate for assessment of the learning curve were also tested (Appendix 1).

Systematic Electronic Bibliographic Database Searching

Eight databases were searched systematically: MEDLINE (1966–March 1999); HealthSTAR (1975 to November 1998); EMBASE (1980–February 1999); Science Citation Index (1981–March 1999); Social Science Citation Index (1981 to March 1999); CINAHL (1982–December 1998); BIOSIS (1985 to March 1999); and the Cochrane Library (1999).

To estimate the number of studies that described the assessment of the learning curve in the body of an article, but which would not have been identified by searching the abstract and title only, we searched the full text of the following databases: MEDLINE Core Biomedical Collection (1993 to August 1998); Biomedical Collection II (1995 to October 1998); and

Biomedical Collection III (1995 to June 1998) (Appendix 1). This covered 46 journals in total.

We searched two electronic databases of ongoing studies: the British National Research Register (Issue 1, 1998) and Current Controlled Trials to January 1999. We also searched the NHS Economic Evaluation Database (NEED) to April 1999.

Hand-searching of Specific Journals

We could not identify any journal for which hand-searching was likely to yield a substantial dividend in extra relevant studies identified. The relevant literature covers too many fields and journals. However, we did identify the *International Journal of Technology Assessment in Health Care* as the place where new techniques to assess learning curves were most likely to be published. Rather than a full hand-search of the journal, a hand-search of all abstracts of full papers was undertaken.

Other Methods of Ascertainment of Studies

We contacted experts in the field, mainly members of the International Society for Health Technology Assessment and biostatisticians, to identify any other relevant studies.

Register of Possible Studies

All possibly relevant reports were electronically imported or manually entered into the software package Reference Manager. Details of the source of each article were added. All electronically derived abstracts and study titles were read by one statistician to assess subject relevance. They were deemed possibly relevant if they described a health technology assessment and also referred to a learning curve. If so, they were assigned keywords in Reference Manager and the full published paper was obtained. The exception was the searching of the full-text version of MEDLINE, where the full published paper was assessed for relevance to learning curves.

Full copies of study reports were assessed for subject relevance, eligibility, and methodologic quality by the statistician using a standard form. The assessor was not blinded to author, institution, or journal.

Inclusion Criteria

To be included in the review, a study had to analyze the learning curve formally by a graph, table, or statistical technique. We categorized the methods of analysis as follows:

- *Descriptive*: No statistical testing was performed, but results were tabulated by experience or shown graphically. The graphical method required one axis to be the case sequence (or grouped case sequence).
- *Split group*: The data were split by experience, and univariate testing of the discrete groups (generally halves or thirds) was performed. The statistical methods used included *t* test, chi-squared test, Mann-Whitney *U* test, and simple ANOVA. Also included in this category are reports that compared experienced with inexperienced surgeons.
- *Univariate (trend)*: These tested for some form of trend by experience in the data. These methods included curve fitting, chi-squared test for trend or repeated measures ANOVA. If the data were split into categories, we required at least three categories with the ordering formally taken into account.
- *Multivariate (split)*: The data were split by experience as in the split group above, and multivariate testing of the groups was performed to adjust for other variables. For example, a study was included in this category if the experience variable had been dichotomized into the first 50 and the second 50 patients, and then included as a potential confounding variable in a regression analysis along with confounders such as age and sex of patient.

- *Multivariate (trend)*: Trend by experience was tested for in the data, but adjustment for possible confounding variables was also included. These methods include logistic regression, and multiple regression with the experience variable treated as either continuous or ordinal.
- *Cumulative sum (CUSUM)*: Trend in experience was measured using the cumulative sum procedure (2). This is a graphical method for identifying trends in data.

Data Abstraction and Analysis

A single statistician abstracted study design, study size, type of technology (minimal access, other surgical, or diagnostic), type of patient, level of learning assessed (operator or institution), number of operators, proportion of operators performing half of the procedures (to see whether one or a few operators dominated the series), type of institution, data source, prior knowledge of outcome before inclusion of patient, type of outcome used to assess learning, and the statistical method used (categorized as above).

A random 10% sample of possible studies was independently assessed by another statistician, with double abstraction of data from those studies meeting the inclusion criteria. A kappa statistic was calculated to measure agreement between the assessors. Any differences of opinion were resolved by discussion.

RESULTS

Literature Search

Of 4,571 abstracts assessed, 559 (12%) were deemed appropriate for further investigation and 272 were later judged on review of the full paper to have included a formal assessment of the learning curve (Table 1). Of the 272 studies, 202 (74%) were identified in MEDLINE. The next largest number was identified from EMBASE, but this is at least in part a function of the order of the searching rather than the coverage of each database. Of the included papers, 39 (15%) were published in non-English language journals. A further 24 (4%) of the 560 assessed full papers mentioned that learning did (or did not) take place in their study, but gave no indication as to how it had been assessed, and hence were excluded. The year of publication is displayed in Figure 1, and shows a progressive rise in number of relevant studies, particularly during the 1990s.

We identified an additional seven papers in the MEDLINE (full-text) database that assessed learning, but which would not have been identified by searching the abstract and

Table 1. Summary of the Bibliographic Searches

Source	Number of abstracts assessed	Number of full papers assessed	Papers included in the review
MEDLINE	736	362	202
EMBASE	588	64	30
Science Citation Index	1235	43	25
MEDLINE (full text)	66	66	7
BIOSIS	629	5	5
CINAHL	28	2	1
HealthSTAR	21	3	1
The Cochrane Library	54	4[a]	0
International Journal of Technology Assessment in Health Care	862	10	1
Social Science Citation Index	352	0	0
Total	4,571	559	272

[a] These were four systematic reviews of technologies that mentioned learning curve effects.

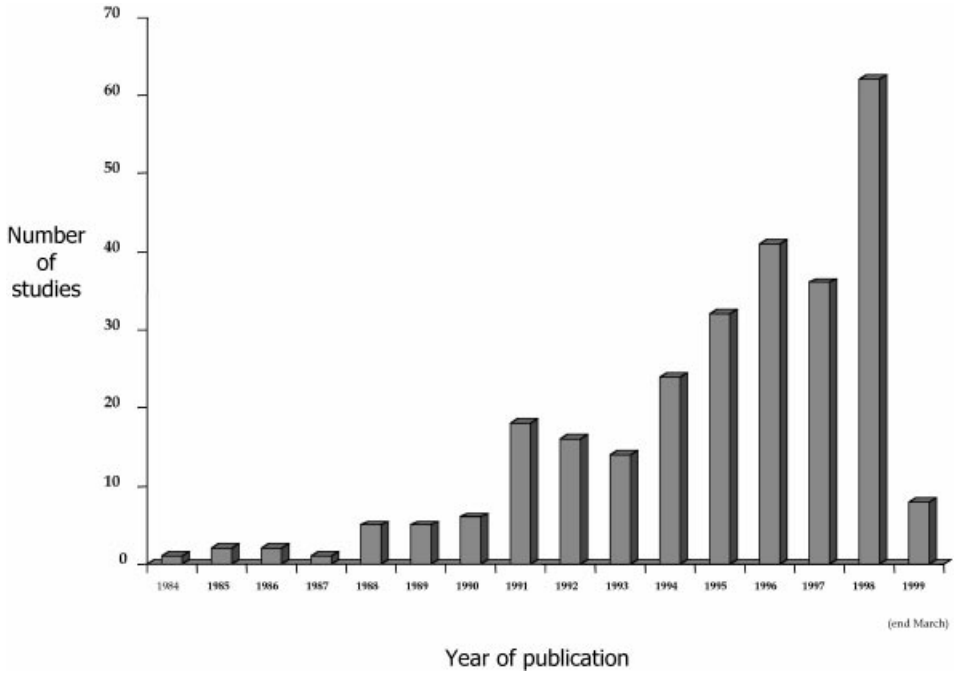

Figure 1. Year of publication of included studies.

title only (Table 1). However, these papers did not contain any further novel techniques. Of the 46 questionnaires sent to experts in the field, 35 (76%) were returned. No extra studies were identified, but additional methods were suggested.

The double assessing of the 10% sample of possible studies showed perfect agreement on the inclusion of papers (kappa = 1) and very good agreement on the methods used (kappa = 0.81). All disagreements were due to omitting a descriptive method when split group methods were also used. This high rate of agreement convinced us that double-assessing of all the papers was not justified. The hand-searching of abstracts from the *International Journal of Technology Assessment in Health Care* identified one additional study but no new method.

Included Studies

Of the 272 included studies, 140 (51%) assessed a surgical minimal access technique such as laparoscopic cholecystectomy, hernia repair, or fundoplication (Table 2). Forty-one percent assessed other treatment procedures such as heart transplantation, and 8% assessed diagnostic technologies such as interpretation of MRI scans. Most of the techniques (96%) were performed on humans.

The majority of the studies (95%) were case series. Only 2% used data collected from a randomized controlled trial. The study sizes varied considerably, but about 40% were less than 100. In 64%, the study addressed the learning curve for a single operator or a single institution only. Approximately half of the studies assessed learning only by individual operators. The remainder assessed learning at the level of the institution (or both).

A few of the studies with more than one operator were dominated by a small number of operators; for example, fewer than 10% of the operators may have performed 50% of the procedures. However, this was unclear in 29% of the studies.

Most of the studies were performed in either tertiary or secondary care centers, but the level of care was unclear for nearly a third. We did not identify any studies in primary care.

Table 2. Study Characteristics

Study characteristic	Number	Percent
Type of procedure		
Laparoscopic	140	51
Other surgical	110	41
Diagnostic	22	8
Assessed on		
Humans	262	96
Animals	6	2
Machines	4	2
Study design		
Case series	259	95
Controlled, nonrandomized	7	3
Randomized controlled	6	2
Study size		
0–50	58	22
51–200	114	41
201–800	59	22
>800	41	15
Number of operators or institutions		
1 only	174	64
2 to 5	43	16
6 to 20	25	9
Over 20	30	11
Level of assessment		
Operators	128	47
Institutions	140	52
Operators and institutions	3	1
Not operators or institutions	1	<1
Proportion of surgeons performing 50% of the procedures		
All	141	52
1% to 10%	7	3
11% to 30%	16	6
31% to 50%	28	10
Unclear	80	29
Level of care of institutions		
Tertiary	98	36
Secondary	66	25
Mixed (tertiary and secondary care)	23	8
Unclear	85	31
Data source		
Prospective	108	40
Retrospective	71	26
Registry	10	4
Unclear	83	31
Prior knowledge of outcome		
Prior knowledge	71	26
No prior knowledge	101	37
Unclear	100	37
Type of outcome used to assess learning		
Intraoperative - continuous	122	45
Intraoperative - dichotomous (not rare)	138	51
Intraoperative - dichotomous (rare)	84	31
Intraoperative - categorical	2	1
Postoperative - continuous	38	14
Postoperative - dichotomous (not rare)	22	8
Postoperative - dichotomous (rare)	15	6
Postoperative - categorical	1	<1

The data were collected prospectively in 40% of the studies and retrospectively in 26%, but this was unclear in 34%. The outcome was known before the analysis of learning began in 26% of the studies and was not known in 37%, but this was unclear in 37%. In some studies data were collected prospectively, but only submitted to a registry or study after the outcome was known.

The types of outcome used to assess learning were mainly intraoperative continuous process variables (45%) such as operation time, and intraoperative dichotomous outcome variables (51%) such as complications. Rare intraoperative events were mentioned in 31% of studies, and rare postoperative events in 6%.

We also examined the interrelationship between some of the study characteristics and the type of technology. The minimal access studies were more likely to have fewer than 50 patients (29%) than studies of other surgical procedures (16%). The type of variables used to assess learning differed between minimal access and other surgical procedures. Minimal access studies more commonly used continuous outcomes than dichotomous outcomes (63% vs. 40%). This pattern was reversed in the other surgical studies where continuous outcomes were less commonly used (30% vs. 59%). The majority of the diagnostic studies used dichotomous outcomes.

Statistical Methods Used

Descriptive. In 119 (44%) studies the data were displayed graphically as a plot of outcome against experience or as a table reporting when results or complications occurred within the series. In all studies in this group, this was done without statistical analysis.

Split Groups. The most common statistical method involved splitting the data into groups by experience—done in 165 (60%) studies. This was usually performed by arbitrary splitting of the series of consecutive cases for individual operators into halves or thirds. The means of the two or three groups were then compared by *t* test or analysis of variance. If these means differed, the authors assumed that learning had taken place. Alternatively, a chi-squared test was used for dichotomous outcomes such as complication rates. Eight (5%) of the splitting studies compared mean operation time between experienced operators with inexperienced operators to test whether the extent of learning differed between the groups.

Univariate Trend. A more sophisticated approach was that of fitting a line to the data by least-squares regression, and this was used in 25 (9%) of the studies (Table 3).

Table 3. Statistical Methods Used in Included Studies

Statistical method	Number	Percent
Descriptive	119	44
Split groups (no test for trend)	165	60
Univariate (trend)	33	12
Curve fitting	25	9
χ^2 test for trend	2	1
Pearson correlation	2	1
Repeated measures ANOVA	3	1
Komolgorov-Smirnoff	1	<1
Multivariate (split–experience dichotomized)	4	1
Logistic regression	3	1
Cox's regression	1	<1
Multivariate (trend–experience continuous)	6	2
Multiple regression	2	1
Logistic regression	3	1
Generalized linear mixed models	1	<1
Cumulative sum	6	2

A linear relationship between experience and outcome was most commonly described (3;19;24;26;27;28;31;32;44;51;53;57). A variety of other curves were used to describe the learning relationship: logarithmic (38;45;48;50), negative exponential (14;49), double negative exponential (55), power form (16), reciprocal (4), quadratic (21;29), and cubic (59). In addition to these 24 papers using least-squares regression, another paper used Monte Carlo simulation to estimate the shape of the learning curve (33).

The correlation between experience and outcome was tested by Spearman's correlation coefficient (35;42), chi-squared test for trend (36;41) or a Komolgorov-Smirnoff test (5). Three studies attempted to model the relationship between experience and outcome using repeated measures ANOVA (7;10;25).

Multivariate (Split Group). A number of multivariate techniques were used. Logistic regression was used in three studies to test whether there was a relationship between a dichotomous outcome and experience (1;6;20). Another study used Cox's regression to look for a learning effect in time-to-event data (18). All these studies adjusted for other confounding factors such as age or sex, but split the experience variable arbitrarily into equal categories.

Multivariate (Trend). The remaining papers kept the experience variable as continuous. Three papers used logistic regression (9;54;56) and two papers used multiple regression (25;58) to adjust for confounding factors before testing for a relationship between experience and operation time. Generalized linear mixed models were used once (30). Multivariate techniques of either type were reported increasingly from 1996.

Cumulative Sum. A cumulative sum technique was used in six studies (4;33;34;39; 43;47).

Split group methods were used less by minimal access studies (51%) than by other surgical studies (67%). Similarly descriptive methods were used more by minimal access studies (53%) than by other surgical studies (37%). The split group method was generally used for assessing the learning curve in diagnostic studies.

DISCUSSION

Systematic Search Strategy

We aimed to describe the "epidemiology" of statistical methods that investigators have used to assess learning in health technology assessment. To avoid bias, a systematic approach was used to identify relevant studies and to extract data. To make the task manageable, the search strategy had to be kept sufficiently specific to avoid highlighting a large number of irrelevant papers. Even limiting this strategy to searching for "learning curve" produced nearly 5,000 abstracts requiring assessment. Exploratory searching using other search terms made clear that the dividend was not worth the resources required. The full-text searching in MEDLINE allowed us to assess how many relevant articles might have been missed as a consequence of basing the search on titles and abstracts only. This search did identify seven (3%) studies that would otherwise have been missed; however, these studies did not provide any statistical methods that had not been identified elsewhere.

We did further searches for other statistical techniques after performing the review of the included studies. After identifying additional statistical methods known to us that could have been used to assess learning in the clinical field, we created a new strategy to search for these, and assessed the abstracts generated for relevance to learning curves. We found no evidence that any of these had been used for this purpose, so we think it unlikely that an important technique has been missed.

Approximately three-quarters of the included studies were identified in MEDLINE. This reflects the ordering of our searches since MEDLINE was searched first. As 70 studies were identified only in databases outside MEDLINE, this confirmed the importance of broader searching.

Use of the term *learning curve* increased over time. During the early 1980s, the term was rare and mainly concerned organ transplantation. The increase since the late 1980s coincided with the introduction of minimally invasive procedures, especially laparoscopic cholecystectomy. However, some 40% of the included studies were concerned with other surgical procedures. Only 8% of included studies concerned diagnostic methods.

Proxies for Learning

The 273 studies use two types of variable to assess learning—measures of patient outcome or quality assurance (13) and measures of clinical process or task efficiency (13). Unfortunately, the patient outcomes used, though acceptable proxies for the goals of health care, tend to be dichotomous rare events like complications or survival and are therefore relatively intractable to statistical analysis. This may be why most studies choose to use continuous process measures; typical examples in surgery are the time to complete an operation and the time that a patient stays in hospital. In minimal access surgery, we found that operation time was more commonly used to assess learning than in other surgical procedures. Although operation time is relatively easy to collect, it is only a weak proxy for learning and does not necessarily relate to proficiency (15;43). As Darzi and colleagues (15) point out, "measuring competence merely by setting time targets for certain procedures is crude and probably unacceptable." Other proxies have been suggested, such as movement of instruments (15) or "near misses" (17), but these too are probably weak proxies for patient outcome and thus learning.

Statistical Methods Used

Our review confirmed that the statistical methods used to assess learning in health technology assessment have almost always been crude. A substantial number of studies have relied upon descriptive data to claim learning without any formal statistical testing.

The most common formal approach was the split group method. Often papers gave no rationale for the cut points, raising concerns about bias caused by data-dependent splitting. Arbitrarily splitting the data into halves was not uncommon. Yet it takes a minimum of three points before one can characterize a trend. Even when splitting suggests that learning has occurred, it is not possible to describe the underlying curve or to identify where particular operators lie on that curve.

A univariate test for trend using curve-fitting procedures was the most commonly used of the more advanced techniques. Papers used a variety of different shapes, but rarely gave a rationale for that selected. A linear relationship was often described, but this could reflect the fact that the series was too short and the operators had not yet reached their final asymptote or plateau.

Multivariate techniques that adjust for a drift in case mix are more robust and potentially useful for investigating trends over time. Unfortunately, the studies we identified have not maximized the potential of these methods. First, some studies dichotomized the experience variable and thus have limitations similar to the split group studies. Second, few studies have attempted to model interoperator differences.

The CUSUM technique has been advocated as the method for monitoring surgical performance (17;52). This technique can be useful for identifying when an operator begins to perform poorly, but it is not so effective for describing interoperator differences. This method has little place within a health technology assessment based on a randomized controlled trial, but it is useful for exploratory analysis.

The assessment of learning curves in diagnostic technologies was not the primary aim of our study, but operators have been compared through receiver-operator characteristic curves (22).

Individual or Institutional Learning

The included studies generally considered learning only within an individual operator or institution. While this approach is useful in looking for learning curves, it suffers from three inherent weaknesses. First, since there is no comparison with other operators or institutions, it is difficult to assess where the operator is on the learning curve. Second, rare complications cannot assess whether there is a relationship between experience and complication rate for one operator. Last, these problems are aggravated by the tendency of single-operator studies to rely on retrospective data collection from medical records, raising concern about the danger of biased abstraction.

It is therefore desirable to obtain prospective data on many operators or institutions. In particular, the creation of data registries for specific technologies could provide a resource for assessing learning curves. To avoid bias, such registries should be prospective and outcomes should not be collected before the patient is registered. However, the continual updating, disseminating, and funding of such registries is difficult (23).

Learning by individuals and by institutions are inextricably linked. Institutional learning adapts processes like those governing referral, patient selection, and aftercare to the circumstances of the new technology. At the same time, individual operators refine their skills in performing the procedure. Any statistical analysis of learning curves should account for this inherent hierarchy.

Randomized Controlled Trials

Nearly all the included studies were case series. Only five (2%) were randomized controlled trials. This could reflect our search strategy. Assessing the learning curve seems more likely to be presented as only a small part of the analysis of a trial, and hence less likely to be mentioned in the abstract. We therefore searched the Cochrane Register of Controlled Trials to find more randomized trials that assessed learning, but found none.

Initial patients receiving a particular technology tend to be either relatively more fit or relatively more sick than those for whom it is later judged to be appropriate (8;13). Within a randomized comparison, such drift in case mix will apply to both groups equally and can be taken into account during analysis.

There are strong arguments that assessment of nonpharmacological technologies should include a pragmatic randomized trial and that this should start as soon as feasible (46). Nevertheless, we recognize that this will not be the only element of assessment. Many assessments will include a prerandomization phase of observational data collection as the technique is developed or refined. Therefore, methods for evaluating learning in these studies are also needed. Such methods would also help to decide whether and when an operator has reached a particular level of competence, and to monitor subsequent performance.

Implications for the Design of Studies

This review has implications for the design of studies, including randomized controlled trials. The experience of the operators should be collected during the study. If this is done, the investigators can look for trends over time throughout the study. As it is unlikely that every patient having a new procedure will be included in a randomized trial, it is also important to record the number of procedures performed between randomized patients.

Implications for the Reporting of Primary Studies

Completed studies need better reporting of the key factors that may be related to the learning curve. As a minimum standard, the number and experience of the operators, the data source, and the level of care should be explicitly mentioned. Our review has shown poor reporting of these factors, causing problems with interpretation and generalization. In particular, an unreported data source implies that one must be cautious about the validity of the study. The level of care could also affect the learning curve; for example, the reporting of results from a tertiary care institution may not be generalizable to secondary care. Finally, the proportion of surgeons performing half of the procedures is important when one wishes to be confident that the aggregated results of a multi-operator study were not influenced by a single operator performing most of the cases.

The difficulties of assessing health technologies with learning curves could be better addressed if rigorous statistical methods were available for measuring and hence adjusting for learning. Randomization could begin as soon as possible consistent with safety and the completion of basic training (12;46), and then continue until well after the learning curve has stabilized. The subsequent analysis would estimate both the point at which the learning curve stabilized and the level of performance achieved (both to within a confidence interval). These two estimates would lead to two distinct but complementary evaluations. The first evaluation would focus on the benefits and costs of introducing the new technology; the second on the benefits and costs of the new technology in steady state. While the second would play the major role in deciding where and when the new technology should be adopted, the first would influence how it should be introduced and what additional training and precautions were needed.

Implications for Future Research

Our review has shown that currently used statistical methods are not sufficiently rigorous. There is a need for methods that can estimate the rate and length of learning together with the final skill level. They should also be capable of exploring and estimating differences between individual operators. We are currently searching for such techniques, notably in fields where learning effects are important, such as psychology and engineering. We are also testing empirically the best of the methods we find on a range of existing data sets. We believe that the development and use of more sophisticated methods will enhance the assessment of health technologies that require learning.

REFERENCES

1. Agachan F, Joo JS, Weiss EG, Wexner SD. Intraoperative laparoscopic complications. Are we getting better? *Dis Colon Rectum*. 1996;39:S14-S19.
2. Altman DG, Royston JP. The hidden effect of time. *Stat Med*. 1988;7:629-637.
3. Archie JP Jr. Learning curve for carotid endarterectomy. *South Med J*. 1988;81:707-710.
4. Atherton DP, O'Sullivan E, Lowe D, Charters P. A ventilation-exchange bougie for fibreoptic intubations with the laryngeal mask airway. *Anaesthesia*. 1996;51:1123-1126.
5. Behrens E, Schramm J, Zentner J, Konig R. Surgical and neurological complications in a series of 708 epilepsy surgery procedures. *Neurosurgery*. 1997;41:1-9.
6. Bennett CL, Stryker SJ, Ferreira MR, Adams J, Beart RW Jr. The learning curve for laparoscopic colorectal surgery. Preliminary results from a prospective analysis of 1194 laparoscopic-assisted colectomies. *Arch Surg*. 1997;132:41-44.
7. Blumenthal PD, Gaffikin L, Affandi B, et al. Training for Norplant implant removal: Assessment of learning curves and competency. *Obstet Gynecol*. 1997;89:174-178.
8. Bouchard S, Barkun AN, Barkun JS, Joseph L. Technology assessment in laparoscopic general surgery and gastrointestinal endoscopy: Science or convenience? *Gastroenterology*. 1996;110:915-925.

9. Bubolz B, Case CL, McKay CA, et al. Learning curve for radiofrequency catheter ablation in pediatrics at a single institution. *Am Heart J*. 1996;131:956-960.
10. Buchman CA, Chen DA, Flannagan P, Wilberger JE, Maroon JC. The learning curve for acoustic tumor surgery. *Laryngoscope*. 1996;106:1406-1411.
11. Buxton MJ. Problems in the economic appraisal of new health technology: The evaluation of heart transplants in the UK. In: Drummond MF, ed. *Economic appraisal of health technology in the European Community*. Oxford: Oxford Medical Publications; 1987:103-118.
12. Chalmers TC. Randomization of the first patient. *Med Clin North Am*. 1975;59:1035-1038.
13. Cuschieri A. Whither minimal access surgery: Tribulations and expectations. *Am J Surg*. 1995;169:9-19.
14. Danford DA, Kugler JD, Deal B, et al. The learning curve for radiofrequency ablation of tachyarrhythmias in pediatric patients. Participating members of the Pediatric Electrophysiology Society. *Am J Cardiol*. 1995;75:587-590.
15. Darzi A, Smith S, Taffinder N. Assessing operative skill. *BMJ*. 1999;318:887-888.
16. Davis Z, Jacobs HK, Zhang M, Thomas C, Castellanos Y. Endoscopic vein harvest for coronary artery bypass grafting: Technique and outcomes. *J Thorac Cardiovasc Surg*. 1998;116:228-235.
17. de Leval MR, Francois K, Bull C, Brawn W, Spiegelhalter D. Analysis of a cluster of surgical failures: Application to a series of neonatal arterial switch operations. *J Thorac Cardiovasc Surg*. 1994;107:914-924.
18. Dunphy BC, Shepherd S, Cooke ID. Impact of the learning curve on term delivery rates following laparoscopic salpingostomy for infertility associated with distal tubal occlusive disease. *Hum Reprod*. 1997;12:1181-1183.
19. Fisher KS, Matteson KM, Hammer MD. Laparoscopic cholecystectomy: The Springfield experience. *Surg Laparosc Endosc*. 1993;3:199-203.
20. Ghosh PK, Choudhary A, Agarwal SK, Husain T. Role of an operative score in mitral reconstruction in dominantly stenotic lesions. *Eur J Cardiothorac Surg*. 1997;11:274-279.
21. Gilchrist BF, Vlessis AA, Kay GA, Swartz K, Dennis D. Open versus laparoscopic cholecystectomy: An initial analysis. *J Laparoendosc Surg*. 1991;1:193-196.
22. Goddard CC, Gilbert RJ, Needham G, Deans HE. Routine receiver operating characteristic analysis in mammography as a measure of radiologists' performance. *Br J Radiol*. 1998;71:1012-1017.
23. Gutzwiller F, Chrzanowski R, Paccaud F. Data bases for the assessment of medical technologies: Examples from Europe. *Int J Technol Assess Health Care*. 1988;4:65-73.
24. Harkki-Siren P, Sjoberg J. Evaluation and the learning curve of the first one hundred laparoscopic hysterectomies. *Acta Obstet Gynecol Scand*. 1995;74:638-641.
25. Heinz G, Kratochwill C, Schmid S, et al. Sinus node dysfunction after orthotopic heart transplantation: The Vienna experience 1987-1993. *Pacing Clin Electrophysiol*. 1994;17:2057-2063.
26. Higashihara E, Baba S, Nakagawa K, et al. Learning curve and conversion to open surgery in cases of laparoscopic adrenalectomy and nephrectomy. *J Urol*. 1998;159:650-653.
27. Horvath KD, Gray D, Benton L, Hill J, Swanstrom LL. Operative outcomes of minimally invasive saphenous vein harvest. *Am J Surg*. 1998;175: 391-395.
28. Hughes GB. The learning curve in stapes surgery. *Laryngoscope*. 1991;101:1280-1284.
29. Johnson C, Roberts JT. Clinical competence in the performance of fiberoptic laryngoscopy and endotracheal intubation: A study of resident instruction. *J Clin Anesth*. 1989;1:344-349.
30. Jowell PS, Baillie J, Branch MS, et al. Quantitative assessment of procedural competence. A prospective study of training in endoscopic retrograde cholangiopancreatography. *Ann Intern Med*. 1996;125:983-989.
31. Kelsey SF, Mullin SM, Detre KM, et al. Effect of investigator experience on percutaneous transluminal coronary angioplasty. *Am J Cardiol*. 1984;53:56C-64C.
32. Kockerling F, Schneider C, Reymond MA, et al. Early results of a prospective multicenter study on 500 consecutive cases of laparoscopic colorectal surgery. *Surg Endosc Ultrasound and Interventional Techniques*. 1998;12:37-41.
33. Konrad C, Schupfer G, Wietlisbach M, Gerber H. Learning manual skills in anesthesiology: Is there a recommended number of cases for anesthetic procedures? *Anesth Analg*. 1998;86:635-639.

34. Kopacz DJ, Neal JM, Pollock JE. The regional anesthesia "learning curve": What is the minimum number of epidural and spinal blocks to reach consistency? *Reg Anesth*. 1996;21:182-190.
35. Laffel GL, Barnett AI, Finkelstein S, Kaye MP. The relation between experience and outcome in heart transplantation. *N Engl J Med*. 1992;327:1220-1225.
36. Lawrence K, McWhinnie D, Goodwin A, et al. Randomised controlled trial of laparoscopic versus open repair of inguinal hernia: Early results. *BMJ*. 1995;311:981-985.
37. Medical Research Council. *Health technology assessment in surgery: The role of the randomised controlled trial*. London: Medical Research Council; 1994.
38. Meehan JJ, Georgeson KE. The learning curve associated with laparoscopic antireflux surgery in infants and children. *J Pediatr Surg*. 1997;32:426-429.
39. Molloy M, Archer SB, Hasselgren PO, Dalton BJ, Bower RH. Cholangiography during laparoscopic cholecystectomy: Cumulative sum analysis of an institutional learning curve. *Gastroenterol*. 1998;114:S0-163.
40. Neugebauer E, Troidl H, Spangenberger W, Dietrich A, Lefering R. Conventional versus laparoscopic cholecystectomy and the randomized controlled trial. Cholecystectomy Study Group. *Br J Surg*. 1991;78:150-154.
41. Ng DT, Rowe NA, Francis IC, et al. Intraoperative complications of 1000 phacoemulsification procedures: A prospective study. *J Cataract Refract Surg*. 1998;24:1390-1395.
42. Ou CS, Beadle E, Presthus J, Smith M. A multicenter review of 839 laparoscopic-assisted vaginal hysterectomies. *J Am Assoc Gynecol Laparosc*. 1994;1:417-422.
43. Parry BR, Williams SM. Competency and the colonoscopist: A learning curve. *Aust NZJ Surg*. 1991;61:419-422.
44. Peters JH, Ellison EC, Innes JT, et al. Safety and efficacy of laparoscopic cholecystectomy. A prospective analysis of 100 initial patients. *Ann Surg*. 1991;213:3-12.
45. Rege RV, Joehl RJ. A learning curve for laparoscopic splenectomy at an academic institution. *J Surg Res*. 1999;81:27-32.
46. Russell I. Evaluating new surgical procedures. *BMJ*. 1995;311:1243-1244.
47. Schlup MT, Williams SM, Barbezat GO. ERCP: A review of technical competency and workload in a small unit. *Gastrointest Endosc*. 1997;46:48-52.
48. See WA, Cooper CS, Fisher RJ. Predictors of laparoscopic complications after formal training in laparoscopic surgery. *JAMA*. 1993;270:2689-2692.
49. Smith DB, Larsson JL. The impact of learning on cost: The case of heart transplantation. *Hospital and Health Services Administration*. 1989;34:85-97.
50. Smith JE, Jackson AP, Hurdley J, Clifton PJ. Learning curves for fibreoptic nasotracheal intubation when using the endoscopic video camera. *Anaesthesia*. 1997;52:101-106.
51. Starkes JL, Payk I, Hodges NJ. Developing a standardized test for the assessment of suturing skill in novice microsurgeons. *Microsurg*. 1998;18:19-22.
52. Steiner SH, Cook RJ, Farewell VT. Monitoring paired binary surgical outcomes using cumulative sum charts. *Stat Med*. 1999;18:69-86.
53. Tai CT, Chen SA, Chiang CE, et al. The effects of accumulated experience on radiofrequency ablation of accessory pathways. *Jpn Heart J*. 1995;36:729-739.
54. Turjman F, Massoud TF, Sayre J, Vinuela F. Predictors of aneurysmal occlusion in the period immediately after endovascular treatment with detachable coils: A multivariate analysis. *Am J Neuroradiol*. 1998;19:1645-1651.
55. Vossen C, Van Ballaer P, Shaw RW, Koninckx PR. Effect of training on endoscopic intracorporeal knot tying. *Hum Reprod*. 1997;12:2658-2663.
56. Wijnberger LD, van der Schouw YT, Christiaens GC. Learning in medicine: Chorionic villus sampling. *Prenat Diagn*. 2000;20:241-246.
57. Witt PD, Wahlen JC, Marsh JL, Grames LM, Pilgram TK. The effect of surgeon experience on velopharyngeal functional outcome following palatoplasty: Is there a learning curve? *Plast Reconstr Surg*. 1998;102:1375-1384.
58. Woods JR, Saywell RMJ, Nyhuis AW, et al. The learning curve and the cost of heart transplantation. *Health Serv Res*. 1992;27:219-238.
59. Yuen PM, Rogers MS. Laparoscopic management of ovarian masses: The initial experience and learning curve. *Aust NZJ Obstet Gynaecol*. 1994;34:191-194.

APPENDIX 1

Literature Search Strategies

The following search terms for the identification of studies related to learning curves were used:
MEDLINE, EMBASE, CINAHL, HealthSTAR: Learning adj4 curve$.tw.
MEDLINE (full text): Learning adj4 curve$.tx.
Science Citation Index, Social Science Citation Index, BIOSIS: Learn* and curve* (in title, abstracts, and keywords).
The Cochrane Library, National Research Register (NRR): Learn* and curve* (in all fields).
NHS Economic Evaluation Database (NEED): Learning curve$ (all fields).
Current Controlled Trials: Learning (any field, any register).
Search terms for statistical methods used in assessing the learning curve: Curve analysis; hierarch* model*; multilevel model*; random effect* model*; generali#ed estimat* equation*; latent curve model*.
Other search terms tested but rejected were: Skill* and (acquir* or acquisit*); learning rate*; ((operator$ or surgeon$) adj4 experience$).tw.; calibrat* and (skill* or learn*).
Key: $ = wildcard; adj(n) = adjacent, within n words either side of the other term; tw = textword, searches in title and abstract; tx = full text, * = wildcard; # = substitutes for one character.