**International Journal of Technology Assessment in Health Care**, **18:1** (2002), 1–10.


# ASSESSING THE LEARNING CURVE EFFECT IN HEALTH TECHNOLOGIES

## *Lessons from the Nonclinical Literature*

**Craig R. Ramsay**
**Sheila A. Wallace**
*University of Aberdeen*

**Paul H. Garthwaite**
*Open University*

**Andrew F. Monk**
**Ian T. Russell**
*University of York*

**Adrian M. Grant**
*University of Aberdeen*

Abstract

**Introduction:** Many health technologies exhibit some form of learning effect, and this represents a barrier to rigorous assessment. It has been shown that the statistical methods used are relatively crude. Methods to describe learning curves in fields outside medicine, for example, psychology and engineering, may be better.
**Methods:** To systematically search non–health technology assessment literature (for example, PsycLit and Econlit databases) to identify novel statistical techniques applied to learning curves.
**Results:** The search retrieved 9,431 abstracts for assessment, of which 18 used a statistical technique for analyzing learning effects that had not previously been identified in the clinical literature. The newly identified methods were combined with those previously used in health technology assessment, and categorized into four groups of increasing complexity: a) exploratory data analysis; b) simple data analysis; c) complex data analysis; and d) generic methods. All the complex structured data techniques for analyzing learning effects were identified in the nonclinical literature, and these emphasized the importance of estimating intra- and interindividual learning effects.
**Conclusion:** A good dividend of more sophisticated methods was obtained by searching in nonclinical fields. These methods now require formal testing on health technology data sets.

**Keywords:** Learning, Clinical competence, Technology assessment, Biomedical, Models, Statistical

This project was funded by the UK 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.

The performance of many health technologies changes over time, especially initially (18). Further technical development (refinement) may continue after clinical introduction, users may improve as they becomes more familiar with a technology, and performance may be enhanced by infrastructure changes, such as better trained assistants or better organized facilities (3;8;13). These "learning curve" changes generally lead to improved performance, and represent a particular challenge to rigorous assessment.

We previously reviewed and appraised the methods by which the issue of the learning curve had been addressed in health technology assessment (HTA) (16). We found that the reporting of studies describing the learning curve effect in health technologies needed improving and that the statistical methods used had been suboptimal (16). There was a need for statistical methods that were appropriate for the structure of the data. Ideally, these methods should aim to estimate three parameters: a) rate of learning; b) baseline (starting) skill level; and c) final skill level (asymptote). They should also allow individual operator differences to be explored and tested. In addition, there were clear problems with measuring learning curve effects using binary events (for example, presence or absence of complications), especially if the events are rare.

We recognized that learning effects are important in other fields, such as psychology and engineering. Examples are the known learning curve effects in manufacturing processes (21;22) and the learning curve associated with cognitive processes research (10). For this reason, we searched other areas aiming to find novel techniques that had not previously been used in HTA. Novel techniques were those that we had not identified in our search of the clinical literature or which improved on an existing technique. The novel technique also had to model or summarize learning curve type data (performance changes). It is the findings of this search that we report here.

## METHODS

### Systematic Electronic Bibliographic Database Searching

We first explored clinical databases for non-HTA use of methods for assessing learning. We then extended searching into other fields, taking advice on the most important databases from experts in the fields. The searches were incremental, in the sense that the product of each search is the extra dividend obtained from that database after excluding duplicates found previously.

To optimize the return on resources available, a specific (i.e., focused) search was developed for each database. Details of the most specific search terms that were chosen and of other less-specific terms that were rejected because they retrieved too many irrelevant studies are obtainable at www.abdn.ac.uk/public_health/hsru/. There were no language or other limits applied to the search strategy.

Twenty-two electronic databases were searched systematically and are shown in Table 1. Many of the clinical databases also include journals from other fields and specialties. For example, MEDLINE includes a number of psychology journals. We searched six clinical databases. Sixteen nonclinical databases were searched for a 10-year period from 1989 onward unless otherwise stated. A description of each database is obtainable at www.abdn.ac.uk/public_health/hsru/.

### Full-Text Electronic Databases Searched Systematically

We electronically searched the full text of selected journal articles to estimate the number of articles that only described the assessment of the learning curve in the body of the report. To do this we used ingentaJournals database, which provides access to full-text versions of a range of English language academic journals produced by BIDS (Bath, UK; over 550

**Table 1.** Summary of the Literature Searches

| Database | Field | No. of abstracts | Possibly relevant papers[a] | Relevant papers |
|---|---|---|---|---|
| *Clinical databases*[b,c] | | | | |
| MEDLINE | Index Medicus | 736 | 5 | 3 |
| EMBASE | Excerpta Medica | 588 | 5 | 1 |
| CINAHL | Nursing and allied | 28 | 0 | 0 |
| HEALTHSTAR | Health research | 21 | 0 | 0 |
| ISI Science Citation Index | Science | 1,235 | 9 | 0 |
| BIOSIS | Biology[d] | 629 | 0 | 0 |
| Clinical total | | 3,237 | 19 | 4 |
| *Nonclinical databases*[c] | | | | |
| RSC | Chemistry | 13 | 0 | 0 |
| ISI Social Science Citation Index | Social science | 352 | 17 | 4 |
| ISI-Arts and Humanities Citation Index | Arts/humanities | 10 | 0 | 0 |
| PsycLit | Psychology | 242 | 17 | 3 |
| IBSS | Economics | 26 | 0 | 0 |
| ISTP | Scientific conference proceedings | 67 | 0 | 0 |
| Ei Compendex/Page One | Engineering | 346 | 3 | 0 |
| SOCIOFILE | Sociology | 11 | 0 | 0 |
| ABI/INFORM | Business | 562 | 2 | 0 |
| ECONLIT | Economics | 50 | 0 | 0 |
| CAB Abstracts | Agriculture[e] | 17 | 0 | 0 |
| INSPEC | Engineering | 255 | 0 | 0 |
| IngentaJournals Online | Many topics | 14 | 0 | 0 |
| Index to theses (GB & Ireland) | Theses | 8 | 0 | 0 |
| Dissertation abstracts | Theses | 147 | 0 | 0 |
| NASA Technical Reports Server (NTRS) | Space/aviation sciences | 353 | 0 | 0 |
| Nonclinical total | | 2,473 | 39 | 7 |
| *Other sources* | | | | |
| Other terms tested in electronic databases[f] | | 3,375[g] | 32[h] | 1[i] |
| Experts in the field | | 61[b] | 8 | 4 |
| Reference lists (of relevant papers + other lists) | | 21 | 2 | 2 |
| Citation indices | | 264[j] | 15 | 0 |
| Total other sources | | 3,721 | 42 | 7 |
| *Total* | | 9,431 | 100 | 18 |

[a] Full papers assessed if technique not previously used in HTA.
[b] Excludes HTA.
[c] Search term: learning curve.
[d] Included clinical and experimental medicine.
[e] Also includes forestry, animal health, and environmental sciences.
[f] Known methods; Binary terms; Skill acquisition; Learning effect; Slips & mistakes; Other terms.
[g] Known methods: 709; Binary terms: 813; Skill acquisition: 503; Learning effect: 266; Slips & mistakes: 245; Other terms: 839.
[h] Known methods: 19; Skill acquisition: 6; Binary terms: 5; Other terms: 2.
[i] From curve analysis.
[j] Numbers of key papers used = 13.

journals covering economics, engineering, mathematics, psychology and other subjects) online. The search was unrestricted.

### Other Methods of Ascertainment of Articles

The reference lists of included articles were followed up to identify further relevant reports. Citation indices were used to identify the subsequent citation in relevant reports. We also contacted a number of experts in the field and asked them to identify any other relevant novel techniques.

We considered handsearching but did not identify a journal for which handsearching was likely to lead to a substantial dividend in terms of extra relevant articles identified.

### Identification of Possibly Relevant Articles

All electronically derived abstracts and article titles were read by one statistician to assess subject relevance. An abstract was deemed possibly relevant if the study was: a) not an HTA; b) the body of the abstract referred to the statistical modeling of a learning curve; and c) we had not previously identified the statistical method used. Full copies of articles were assessed for subject relevance, eligibility, and methodologic quality by a subject expert. The assessor was not blinded to author, institution, or journal.

### Inclusion Criteria

To be included in the review, a study had to assess a learning curve formally using a novel statistical technique. We categorized the statistical techniques into four subgroups of increasing complexity:

1. *Exploratory data analysis*: techniques that do not estimate statistical parameters or test hypotheses. These included graphical displaying of the data or creating cumulative sums (CUSUM) of consecutive cases (1).
2. *Simple series data analysis*: techniques that use data collected on a single operator or summarized over many operators. For example, a) a *t* test comparing the mean operation time for the first 50 cases with the mean time for the next 50; b) a study of 10 operators performing 20 procedures each, in which the data were analyzed as the average performance of the 10 operators; and c) fitting the best shape of curve to each individual operator's performance and describing the various shapes that these curves took.
3. *Complex data structure analysis*: techniques that use data collected on many individuals and measure both differences between individuals and the overall pattern of learning.
4. *Generic techniques*: techniques that can be applied to both simple series data and complex data structures.

### Double Assessment

Each included study was independently assessed by a second statistician. Any differences of opinion were resolved by discussion.

## RESULTS

### Literature Search

The numbers of possibly relevant abstracts generated by the systematic searches are shown in Table 1. Of 9,431 abstracts assessed, 100 (1%) were deemed appropriate for further investigation. Of these, 18 were later judged on review of the full paper to describe a novel technique or to make a significant addition to a previously recognized technique. The

**Table 2.** Shapes of Learning Curves

| Type of curve | Equation[a] |
|---|---|
| *Curves previously used in HTA* | |
| Linear | $y = a + bX$ |
| Quadratic | $y = a + bX + cX^2$ |
| Cubic | $y = a + bX + cX^2 + dX^3$ |
| Power law (asymptote at zero) | $y = bX^{-c}$ |
| Reciprocal | $y = a + \frac{b}{X}$ |
| Exponential | $y = ae^{-bX}$ |
| Double exponential | $y = ae^{-bX} + ce^{-dX}$ |
| Logarithmic | $y = a \ln X + b$ |
| *Potentially useful curves identified outside HTA* | |
| Power law (with nonzero asymptote) | $y = a + bX^{-c}$ |
| Log-log linear model | $\ln y = a(\ln(X+1))^b$ |
| Log–linear curve | $\ln y = aX^b$ |
| Logistic curve | $y = \frac{a}{1 + be^{-cX}}$ |
| Weibull curve | $y = a - be^{-cX^d}$ |
| Exponential difference equation | $\xi(X) = a - [a - \xi(X-1)]e^{-b}$ |
| Cumulative performance curves | $y = \frac{k}{X}\sum_{X=1}^{N} X^{-s}$ |
| Exponential cumulative performance curves | $y = \frac{(a^N - 1)[b - c]}{(a-1)N} + c$ |

[a] Y denotes the continuous outcome (for example, time), and $X$ denotes the case sequence number (for example, $X = 1$ is the first procedure, $X = 2$ is the second procedure, and so on).

dividend from each of the searches is also shown in Table 1. Of the 18 included studies, four were identified in Social Science Citation Index, and three each from PsycLit and MEDLINE.

Fifty (82%) of the 61 questionnaires sent to experts in the fields were returned. Four additional methodologies were suggested. The reference lists of the selected papers produced two other novel techniques.

## Included Techniques

The 18 included papers were categorized in two ways: a) shapes of learning curves (Table 2); and b) statistical techniques (Table 3). The bottom part of Table 2 displays the eight new shapes of curve identified (with the previously identified curves on top). The additional curves came from psychology (9;10;12;15), manufacturing (2;5), and aviation (19). The most widely cited shape of learning curve across all fields was the power law (of practice). All curves had a similar basic shape that decreased to an asymptote. The logistic curve was identified in one paper only and measured the error rates of pilots in a simulator (19). The curve was also discussed in the psychology literature as an *implausible* learning curve shape because no learning mechanism could explain the point of inflection (15).

Table 3 shows the dividend in statistical techniques from searching the non-HTA literature. The new methods were potentially useful for assessing learning curves using complex structured data that had not been identified in the HTA literature. The following is a brief outline of the more complex methods and the studies, where they were used.

***Principal Components Analysis.*** Principal components analysis was used in one study measuring active avoidance learning in 24 rats with streptozotocin diabetes and compared these with 27 control rats over 100 consecutive trials (23). The average final success rate (asymptote), latency (the period from initial point to 50% success level), and rate (number of trials to reach asymptote) were measured. These measurements were then used in a principal components analysis to search for subgroups of rats with similar learning characteristics. This method estimates groups of individuals that have common responses

**Table 3.** Techniques Used to Detect the Learning Curve

| Technique |
|---|
| *Techniques previously used in HTA* |
| *Exploratory data analysis* |
| Graph |
| CUSUM techniques |
| *Techniques for simple series data* |
| T-test, one way ANOVA |
| Chi-squared test (for trend) |
| Repeated measures ANOVA[a] |
| Curve fitting |
| Multiple regression[a] |
| Logistic regression[a] |
| *Techniques for complex structured data* |
| None identified |
| *Techniques that can be applied to both simple and complex data types* |
| None identified |
| *Potentially useful techniques identified outside HTA* |
| *Exploratory data analysis* |
| None identified |
| *Techniques for simple series data* |
| Curve fitting |
| *Techniques for complex structured data* |
| Principal components analysis |
| Two-stage estimation of learning rates |
| Generalized estimating equations |
| Multilevel models |
| Latent curve models |
| Time series models (ARIMA) |
| Stochastic parameter models |
| *Techniques that can be applied to both simple and complex data types* |
| Generalized linear models |

[a] These techniques are special cases of generalized linear models.

in the data, and therefore could be used to compare characteristics of experienced and inexperienced operators; however, a major weakness of this method is that it cannot be used to measure the rate and asymptote of learning of operators.

***Two-Stage Estimation of Learning Rates.*** Two-stage estimation of learning rates was described in a study involving 115 students performing a simulated air traffic control task on 18 consecutive trials (9). The simulation had three task components: accepting planes into the airspace, moving planes in a three-level hold pattern, and landing planes on appropriate runways. The number of correct landings per trial was the outcome of interest, and individual learning curves were estimated using a negative exponential curve. This is a relatively simple procedure that can be applied using most standard statistical packages. It could be applied to multiple operator HTA studies and would provide estimates of the rate and asymptote of learning as well as an estimate of the between-operator correlation.

***Generalized Estimating Equations.*** One respondent to our questionnaire recommended generalized estimating equations as a method for investigating learning curves. Generalized estimating equations allow the correlation of outcomes within an individual to be estimated and accounted for in any subsequent analyses (6;24). This procedure is similar to the two-stage estimation procedure above and will provide the same amount of information, but is strengthened by the fact that the estimates are obtained iteratively, not just in two stages. An additional advantage of this method is that dichotomous outcomes (such as complications) can be modeled.

***Multilevel Models.*** Multilevel models have increasingly been used in clinical research (11;17), and involve partitioning the variability between and within hierarchies in the data set. Learning curve data have an inherent hierarchy: individuals performing many procedures. This is a promising method for analyzing learning curve effects in HTA, since it will provide estimates of the correlation between operators within and between institutions while simultaneously modeling the effect of the individual operator learning curve. The possible disadvantage is that dichotomous outcomes are difficult to model using this method.

***Latent Curve Models/Stochastic Parameter Models/ARIMA Models.*** Both latent curve models and stochastic parameter models are specialized cases of structural equation models (20). In both cases, factors are calculated for each individual's asymptote, total learning (increase from first to last trial), and rate of learning. The complex structured data techniques identified are completed by time series modeling using autoregressive, integrated moving averages models. A comparison of these three methods was performed using data from a study that measured 137 U.S. Air Force personnel performing an air traffic control task on six occasions (4). Again, the number of correct landings per trial was the outcome of interest. All three of these methods will provide information on the rate, asymptote, and differences between operators. The major problem is that the models appear to perform better if the data are very short series on many individuals. Therefore, the applicability of these methods to clinical trials where, for example, there may be few surgeons performing many operations, is unclear.

***Generalized Linear Models.*** Generalized linear models were included as a method that could be applied to both complex and simple structured data. Multiple and logistic regression are generalized linear models that have been used in the clinical literature to assess learning curves of individual operators. They have not, however, been used to assess differences between operators (although they could have been).

## DISCUSSION

The primary aim of this study was to identify novel statistical techniques that could be used to assess the learning curve effect in HTA by searching the non-HTA literature. We searched a range of potential sources systematically. We used an incremental approach, based on formal assessment of full papers whose abstracts suggested that the statistical techniques had not been identified previously.

The resources required for the searches were considerable. Even limiting searches to simple search terms identified approximately 10,000 abstracts that required assessment. When no relevant papers were identified in a database, it may not mean that there was nothing relevant in that database, but rather that our searches were too limited. We identified 18 papers that used a novel technique. It is possible, however, that other papers may have been missed because of poorly documented details of the methodology in the abstract. Contacting experts in the field proved one of the most fruitful sources of relevant studies. Four of the 18 relevant papers were identified by this method. We recommend that reviews of research methods should include this approach.

There is an important distinction between methods for *identifying* a learning effect and those for *measuring* (characterizing) a learning effect. The exploratory methods—*t* tests, ANOVA, and chi-squared tests—in Table 3 could be used to identify if there was a change in operator outcome with experience. However, these methods cannot measure or quantify the rate or asymptote of learning. In contrast, fitting a particular shape of learning curve to the data can be used to measure a trend with experience. The more advanced modeling techniques use the "best" shape of individual curves to explain observed differences between operators. An extra eight shapes of curves were identified. In contrast to the HTA literature, the non-HTA literature tended to give rational reasons for the shape of curves used; for example, Newell and Rosenbloom (15) give a "chunking" theory of learning to explain a power curve–type relationship. Further work is required to describe the relationship between the various curves and underlying behavioral models. The HTA literature tended to fit a curve without explanation for the choice of shape or comparison of different shapes. We recommend that any analyses of health technologies using simple series data should consider the likely shape of the learning curve *a priori*, bearing in mind that there are a number of different types of learning curves. The find that the power curve relationship has been used across many fields suggests that the power curve should be the standard against which all other curves are measured. All of the shapes in Table 2 may be assessed using curve-fitting techniques that are available in most statistical packages.

We identified nine novel statistical techniques for measuring learning curves. The techniques used for complex structured data were only identified in the non-HTA literature (Table 3). There are clear advantages in using most of these: a measure of how an individual operator performs is obtained together with a measure of how the operator is performing in relation to the other operators in the study. This enables the investigators to explore the influence of each operator in a study. These methods also use more of the data and are statistically more powerful. In addition, the methods can also incorporate other covariates such as case-mix variables. A limitation is that many of the methods perform better with certain types of data. The most promising technique for learning curve data in HTA may be multilevel modeling because learning curve data in HTA have a hierarchy (patients within operators within institutions), but this requires further investigation. Another limitation is the availability of software to apply many of the methods. The more sophisticated the method becomes, the more difficult it is to find a standard statistical package that can perform it. For example, multilevel models are available through MLWin software (14).

Thus, there is a hierarchy of statistical methods that can be used to analyze learning curves. We are not advocating that all learning curve analyses should use the most complex methods, rather they should employ the simplest method that can answer the questions being posed. The methods used should be parsimonious; that is, they should not use more parameters than are necessary. For example, if a Weibull curve and exponential curve produced similar results, then the exponential curve would be preferred since it only requires two parameters to be estimated rather than the four required by the Weibull.

## POLICY IMPLICATIONS

### Implications for Conduct of Clinical Trials

In the past there has been controversy over the timing of the assessment of those technologies that need some form of learning before they are at their most effective or efficient. Early assessment is open to the criticism that the new technology is at a disadvantage because the operators are not fully proficient. Late assessment runs the risk that the operators will draw their own conclusions about the advantages and disadvantages of the new technology. In Buxton's words:

> It's always too early until, unfortunately, it's suddenly too late. (7)

The portfolio of statistical methods described by this review opens the door to another approach. Randomization could begin as soon as possible, consistent with safety and the completion of basic training, 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 are needed.

## Implications for Further Research

We have found that there are a number of more sophisticated statistical methods that could be used to model the learning curve effect during HTA. The relative performance of these methods requires assessment before general recommendations can be made.

### REFERENCES


1. Altman DG, Royston JP. The hidden effect of time. *Stat Med.* 1988;7:629-637.
2. Bailey C, Mcintyre E. Some evidence on the nature of relearning curves. *Accounting Review*. 1992;67:368-378.
3. Bouchard S, Barkun AN, Barkun JS, Joseph L. Technology assessment in laparoscopic general surgery and gastrointestinal endoscopy: Science or convenience? *Gastroenterolgy.* 1996;110:915-925.
4. Browne MW, Du-Toit S-HC. Models for learning data. In Collins LM, Horn JL, et al., eds. *Best methods for the analysis of change*: *Recent advances*, *unanswered questions*, *future directions*. Washington, DC: American Psychological Association; 1991:47-68.
5. Buck J, Cheng S. Instructions and feedback effects on speed and accuracy with different learning-curve models. *IIE Transactions*. 1993;25:34-47.
6. Burton P, Gurrin L, Sly P. Extending the simple linear regression model to account for correlated responses: An introduction to generalized estimating equations and multilevel mixed modelling. *Stat Med.* 1998;17:1261-1291.
7. 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.
8. Cuschieri A. Whither minimal access surgery: Tribulations and expectations. *Am J Surg*. 1995;169:9-19.
9. Eyring JD, Johnson DS, Francis DJ. A cross-level units-of-analysis approach to individual differences in skill acquisition. *J Appl Psychol.* 1993;78:805-814.
10. Fitts PM, Posner MI. Human performance. Belmont, CA: Brooks Cole; 1969.
11. Goldstein H, Healy MJR, Rasbash J. Multilevel time series models with applications to repeated measures data. *Stat Med.* 1994;13:1643-1655.
12. Logan GD. Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *J Exp Psychol Learn Mem Cogn.* 1992;18:883-914.
13. Mowatt G, Bower DJ, Brebner JA, et al. When is the “right” time to initiate an assessment of a health technology? *Int J Technol Assess Health Care*. 1998;14:372-386.
14. Multilevel Models Project. *A user’s guide to MLwiN*. 1998.
15. Newell A, Rosenbloom P. Mechanisms of skill acquisition and the law of practice. In Anderson JR, ed. *Cognitive skills and their acquisition*. Hillsdale, NY: Lawrence Erlbaum Associates; 1981.
16. Ramsay CR, Grant AM, Wallace SW, et al. Assessment of the learning curve in health technologies: A systematic review. *Int J Technol Assess Health Care* 2000;16:1095-1108.
17. Rice N, Leyland A. Multilevel models: Applications to health data. *Journal of Health Services Research Policy*. 1996;1:154-164.
18. Russell I. Evaluating new surgical procedures. *BMJ*. 1995;311:1243-1244.
19. Spears WD. Measurement of learning and transfer through curve fitting. *Human Factors*. 1985;27:251-266.

20. Ullman JB. Structural equation modeling. In Tabachnick BG, Fidell LS, eds. *Using multivariate statistics*. New York: HarperCollins College Publishers; 1996:709-811.
21. Wright TP. Factors affecting the cost of airplanes. *Journal of Aeronautical Science*. 1936;3:122-136.
22. Yelle LE. Estimating learning curves for potential products. *Industrial Marketing Management*. 1976;5:147-154.
23. Zaraiskaya YI, Aleksandrova EA, Lukashev AO, Shvyrkova NA. Features of active avoidance learning in rats with streptozotocin diabetes. *Neurosci Behav Physiol*. 1994;24:167-169.
24. Ziegler A, Kastner C, Blettner M. The generalised estimating equations: An annotated bibliography. *Biometrical Journal*. 1998;40:115-139.