

Analytical studies: a framework for quality improvement design and analysis

Lloyd P Provost

Institute for Healthcare Improvement, Associates in Process Improvement, Austin, Texas, USA

Correspondence to

Lloyd P Provost, Institute for Healthcare Improvement, Associates in Process Improvement, 115 East 5th Street, Suite 300, Austin, TX 78701, USA; lprovost@apiweb.org

Accepted 18 January 2011

ABSTRACT

Conducting studies for learning is fundamental to improvement. Deming emphasised that the reason for conducting a study is to provide a basis for action on the system of interest. He classified studies into two types depending on the intended target for action. An enumerative study is one in which action will be taken on the universe that was studied. An analytical study is one in which action will be taken on a cause system to improve the future performance of the system of interest. The aim of an enumerative study is estimation, while an analytical study focuses on prediction. Because of the temporal nature of improvement, the theory and methods for analytical studies are a critical component of the science of improvement.

INTRODUCTION: ENUMERATIVE AND ANALYTICAL STUDIES

Designing studies that make it possible to learn from experience and take action to improve future performance is an essential element of quality improvement. These studies use the now traditional theory established through the work of Fisher,¹ Cox,² Campbell and Stanley,³ and others that is widely used in biomedicine research. These designs are used to discover new phenomena that lead to hypothesis generation, and to explore causal mechanisms,⁴ as well as to evaluate efficacy and effectiveness. They include observational, retrospective, prospective, pre-experimental, quasiexperimental, blocking, factorial and time-series designs.

In addition to these classifications of studies, Deming⁵ defined a distinction between analytical and enumerative studies which has proven to be fundamental to the science of improvement. Deming based his insight on the distinction between these two approaches that Walter Shewhart had made in 1939 as he helped develop measurement

strategies for the then-emerging science of 'quality control.'⁶ The difference between the two concepts lies in the extrapolation of the results that is intended, and in the target for action based on the inferences that are drawn.

A useful way to appreciate that difference is to contrast the inferences that can be made about the water sampled from two different natural sources (*figure 1*). The enumerative approach is like the study of water from a pond. Because conditions in the bounded universe of the pond are essentially static over time, analyses of random samples taken from the pond at a given time can be used to estimate the makeup of the entire pond. Statistical methods, such as hypothesis testing and CIs, can be used to make decisions and define the precision of the estimates.

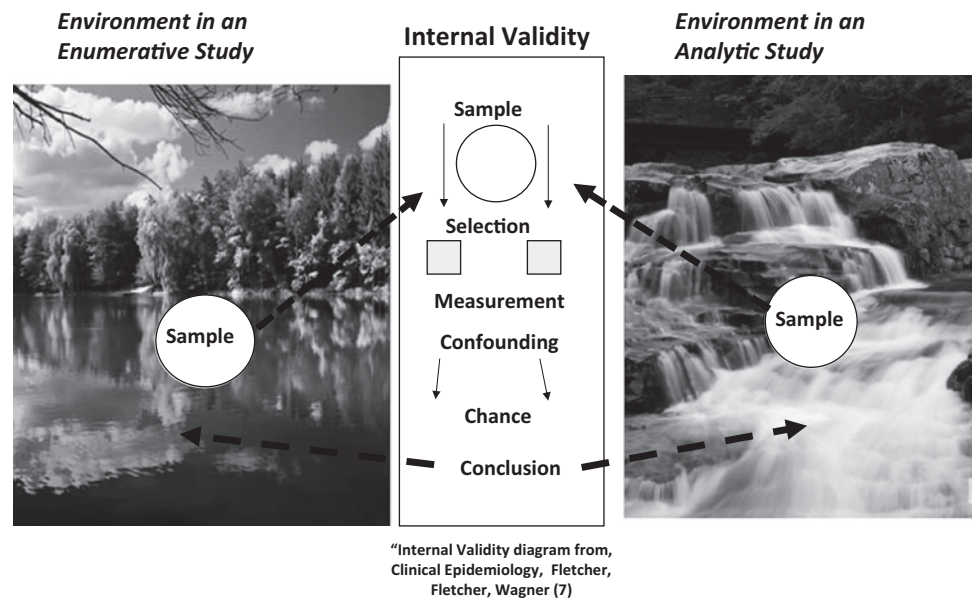
The analytical approach, in contrast, is like the study of water from a river. The river is constantly moving, and its physical properties are changing (eg, due to snow melt, changes in rainfall, dumping of pollutants). The properties of water in a sample from the river at any given time may not describe the river after the samples are taken and analysed. In fact, without repeated sampling over time, it is difficult to make predictions about water quality, since the river will not be the same river in the future as it was at the time of the sampling.

Deming first discussed these concepts in a 1942 paper,⁸ as well as in his 1950 textbook,⁹ and in a 1975 paper used the enumerative/analytical terminology to characterise specific study designs.⁵ While most books on experimental design describe methods for the design and analysis of enumerative studies, Moen *et al*¹⁰ describe methods for designing and learning from analytical studies. These methods are graphical and focus on prediction of future performance. The concept of analytical



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://qualitysafety.bmj.com/site/about/unlocked.xhtml>

Figure 1 Environment in enumerative and analytical study. Internal validity diagram from Fletcher *et al.*⁷



studies became a key element in Deming's 'system of profound knowledge' that serves as the intellectual foundation for improvement science.¹¹ The knowledge framework for the science of improvement, which combines elements of psychology, the Shewhart view of variation, the concept of systems, and the theory of knowledge, informs a number of key principles for the design and analysis of improvement studies:

- ▶ Knowledge about improvement begins and ends in experimental data but does not end in the data in which it begins.
- ▶ Observations, by themselves, do not constitute knowledge.
- ▶ Prediction requires theory regarding mechanisms of change and understanding of context.
- ▶ Random sampling from a population or universe (assumed by most statistical methods) is not possible when the population of interest is in the future.
- ▶ The conditions during studies for improvement will be different from the conditions under which the results will be used. The major source of uncertainty concerning their use is the difficulty of extrapolating study results to different contexts and under different conditions in the future.
- ▶ The wider the range of conditions included in an improvement study, the greater the degree of belief in the validity and generalisation of the conclusions.

The classification of studies into enumerative and analytical categories depends on the intended target for action as the result of the study:

- ▶ Enumerative studies assume that when actions are taken as the result of a study, they will be taken on the material in the study population or 'frame' that was sampled.

More specifically, the study universe in an enumerative study is the bounded group of items (eg, patients, clinics, providers, etc) possessing certain properties of interest. The universe is defined by a frame, a list of identifiable, tangible units that may be sampled and studied. Random selection methods are assumed in the statistical methods used for estimation, decision-making and drawing inferences in enumerative studies. Their aim is estimation about some aspect of the frame (such as a description, comparison or the existence of a cause–effect relationship) and the resulting actions taken on this particular frame. One feature of an enumerative study is that a 100% sample of the frame provides the complete answer to the questions posed by the study (given the methods of investigation and measurement). Statistical methods such as hypothesis tests, CIs and probability statements are appropriate to analyse and report data from enumerative studies. Estimating the infection rate in an intensive care unit for the last month is an example of a simple enumerative study.

- ▶ Analytical studies assume that the actions taken as a result of the study will be on the process or causal system that produced the frame studied, rather than the initial frame itself. The aim is to improve future performance.

In contrast to enumerative studies, an analytical study accepts as a given that when actions are taken on a system based on the results of a study, the conditions in that system will inevitably have changed. The aim of an analytical study is to enable prediction about how a change in a system will affect that system's future performance, or prediction as to which plans or strategies for future action on the system will be superior. For example, the task may be to choose among several different treatments for future patients, methods of

collecting information or procedures for cleaning an operating room. Because the population of interest is open and continually shifts over time, random samples from that population cannot be obtained in analytical studies, and traditional statistical methods are therefore not useful. Rather, graphical methods of analysis and summary of the repeated samples reveal the trajectory of system behaviour over time, making it possible to predict future behaviour. Use of a Shewhart control chart to monitor and create learning to reduce infection rates in an intensive care unit is an example of a simple analytical study.

The following scenarios give examples to clarify the nature of these two types of studies.

Scenario 1: enumerative study—observation

To estimate how many days it takes new patients to see all primary care physicians contracted with a health plan, a researcher selected a random sample of 150 such physicians from the current active list and called each of their offices to schedule an appointment. The time to the next available appointment ranged from 0 to 180 days, with a mean of 38 days (95% CI 35.6 to 39.6).

Comment

This is an enumerative study, since results are intended to be used to estimate the waiting time for appointments with the plan's current population of primary care physicians.

Scenario 2: enumerative study—hypothesis generation

The researcher in scenario 1 noted that on occasion, she was offered an earlier visit with a nurse practitioner (NP) who worked with the physician being called. Additional information revealed that 20 of the 150 physicians in the study worked with one or more NPs. The next available appointment for the 130 physicians without an NP averaged 41 days (95% CI 39 to 43 days) and was 18 days (95% CI 18 to 26 days) for the 20 practices with NPs, a difference of 23 days (a 56% shorter mean waiting time).

Comment

This subgroup analysis suggested that the involvement of NPs helps to shorten waiting times, although it does not establish a cause–effect relationship, that is, it was a ‘hypothesis-generating’ study. In any event, this was clearly an enumerative study, since its results were to understand the impact of NPs on waiting times in the particular population of practices. Its results suggested that NPs might influence waiting times, but only for practices in this health plan during the time of the study. The study treated the conditions in the health plan as static, like those in a pond.

Scenario 3: enumerative study—comparison

To find out if administrative changes in a health plan had increased member satisfaction in access to care, the customer service manager replicated a phone survey he had conducted a year previously, using a random sample of 300 members. The percentage of patients who were satisfied with access had increased from 48.7% to 60.7% (Fisher exact test, $p < 0.004$).

Comment

This enumerative comparison study was used to estimate the impact of the improvement work during the last year on the members in the plan. Attributing the increase in satisfaction to the improvement work assumes that other conditions in the study frame were static.

Scenario 4: analytical study—learning with a Shewhart chart

Each primary care clinic in a health plan reported its ‘time until the third available appointment’ twice a month, which allowed the quality manager to plot the mean waiting time for all of the clinics on Shewhart charts. Waiting times had been stable for a 12-month period through August, but the manager then noted a special cause (increase in waiting time) in September. On stratifying the data by region, she found that the special cause resulted from increases in waiting time in the Northeast region. Discussion with the regional manager revealed a shortage of primary care physicians in this region, which was predicted to become worse in the next quarter. Making some temporary assignments and increasing physician recruiting efforts resulted in stabilisation of this measure.

Comment

Documenting common and special cause variation in measures of interest through the use of Shewhart charts and run charts based on judgement samples is probably the simplest and commonest type of analytical study in healthcare. Such charts, when stable, provide a rational basis for predicting future performance.

Scenario 5: analytical study—establishing a cause–effect relationship

The researcher mentioned in scenarios 1 and 2 planned a study to test the existence of a cause–effect relationship between the inclusion of NPs in primary care offices and waiting time for new patient appointments. The variation in patient characteristics in this health plan appeared to be great enough to make the study results useful to other organisations. For the study, she recruited about 100 of the plan's practices that currently did not use NPs, and obtained funding to facilitate hiring NPs in up to 50 of those practices.

The researcher first explored the theories on mechanisms by which the incorporation of NPs into primary care clinics could reduce waiting times. Using important contextual variables relevant to these mechanisms (practice size, complexity, use of information technology and urban vs rural location), she then developed a randomised block, time-series study design. The study had the power to detect an effect of a mean waiting time of 5 days or more overall, and 10 days for the major subgroups defined by levels of the contextual variables. Since the baseline waiting time for appointments varied substantially across practices, she used the baseline as a covariate.

After completing the study, she analysed data from baseline and postintervention periods using stratified run charts and Shewhart charts, including the raw measures and measures adjusted for important covariates and effects of contextual variables. Overall waiting times decreased 12 days more in practices that included NPs than they did in control practices. Importantly, the subgroup analyses according to contextual variables revealed conditions under which the use of NPs would not be predicted to lead to reductions in waiting times. For example, practices with short baseline waiting times showed little or no improvement by employing NPs. She published the results in a leading health research journal.

Comment

This was an analytical study because the intent was to apply the learning from the study to future staffing plans in the health plan. She also published the study, so its results would be useful to primary care practices outside the health plan.

Scenario 6: analytical study—implementing improvement

The quality-improvement manager in another health plan wanted to expand the use of NPs in the plan's primary care practices, because published research had shown a reduction in waiting times for practices with NPs. Two practices in his plan already employed NPs. In one of these practices, Shewhart charts of waiting time by month showed a stable process averaging 10 days during the last 2 years. Waiting time averaged less than 7 days in the second practice, but a period when one of the physicians left the practice was associated with special causes.

The quality manager created a collaborative among the plan's primary care practices to learn how to optimise the use of NPs. Physicians in the two sites that employed NPs served as subject matter experts for the collaborative. In addition to making NPs part of their care teams, participating practices monitored appointment supply and demand, and tested other changes

designed to optimise response to patient needs. Thirty sites in the plan voluntarily joined the collaborative and hired NPs. After 6 months, Shewhart charts indicated that waiting times in 25 of the 30 sites had been reduced to less than 7 days. Because waiting times in these practices had been stable over a considerable period of time, the manager predicted that future patients would continue to experience reduced times for appointments. The quality manager began to focus on a follow-up collaborative among the backlog of 70 practices that wanted to join.

Comment

This project was clearly an analytical study, since its aim was specifically to improve future waiting-time performance for participating sites and other primary care offices in the plan. Moreover, it focused on learning about the mechanisms through which contextual factors affect the impact of NPs on primary care office functions, under practice conditions that (like those in a river) will inevitably change over time.

DISCUSSION

Statistical theory in enumerative studies is used to describe the precision of estimates and the validity of hypotheses for the population studied. But since these statistical methods provide no support for extrapolation of the results outside the population that was studied, the subject experts must rely on their understanding of the mechanisms in place to extend results outside the population.

In analytical studies, the standard error of a statistic does not address the most important source of uncertainty, namely, the change in study conditions in the future. Although analytical studies need to take into account the uncertainty due to sampling, as in enumerative studies, the attributes of the study design and analysis of the data primarily deal with the uncertainty resulting from extrapolation to the future (generalisation to the conditions in future time periods). The methods used in analytical studies encourage the exploration of mechanisms through multifactor designs, contextual variables introduced through blocking and replication over time.

Prior stability of a system (as observed in graphic displays of repeated sampling over time, according to Shewhart's methods) increases belief in the results of an analytical study, but stable processes in the past do not guarantee constant system behaviour in the future. The next data point from the future is the most important on a graph of performance. Extrapolation of system behaviour to future times therefore still depends on input from subject experts who are familiar with mechanisms of the

system of interest, as well as the important contextual issues. Generalisation is inherently difficult in all studies because ‘whereas the problems of internal validity are solvable within the limits of the logic of probability statistics, the problems of external validity are not logically solvable in any neat, conclusive way’³ (p. 17).

The diverse activities commonly referred to as healthcare improvement¹² are all designed to change the behaviour of systems over time, as reflected in the principle that ‘not all change is improvement, but all improvement is change.’ The conditions in the unbounded systems into which improvement interventions are introduced will therefore be different in the future from those in effect at the time the intervention is studied. Since the results of improvement studies are used to predict future system behaviour, such studies clearly belong to the Deming category of analytical studies. Quality improvement studies therefore need to incorporate repeated measurements over time, as well as testing under a wide range of conditions (2, 3 and 10). The ‘gold standard’ of analytical studies is satisfactory prediction over time.

CONCLUSIONS AND RECOMMENDATIONS

In light of these considerations, some important principles for drawing inferences from improvement studies include¹⁰:

1. The analysis of data, interpretation of that analysis and actions taken as a result of the study should be closely tied to the current knowledge of experts about mechanisms of change in the relevant area. They can often use the study to discover, understand and evaluate the underlying mechanisms.
2. The conditions of the study will be different from the future conditions under which the results will be used. Assessment by experts of the magnitude of this difference and its potential impact on future events should be an integral part of the interpretation of the results of the intervention.
3. Methods for the analysis of data should be almost exclusively graphical, with the aim of partitioning the data visually among the sources of variation present in

the study. In reporting the results of an improvement project, authors should consider the following general guidelines for the analysis:

- ▶ Show all the data before aggregation or summary.
- ▶ Plot the outcome data in the order in which the tests of change were conducted and annotate with information on the interventions.
- ▶ Use graphical displays to assess how much of the variation in the data can be explained by factors that were deliberately changed.
- ▶ Rearrange and subgroup the data to study other sources of variation (background and contextual variables).
- ▶ Summarise the results of the study with appropriate graphical displays.

Because these principles reflect the fundamental nature of improvement—taking action to change performance, over time, and under changing conditions—their application helps to bring clarity and rigour to improvement science.

Acknowledgements The author is grateful to F Davidoff and P Batalden for their input to earlier versions of this paper.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Fisher RA. *Design of Experiment*. 8th edn. New York: Hafner Publishing Company, 1966.
2. Cox DR. *Planning of Experiments*. New York: John Wiley and Sons, 1958.
3. Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin Company, 1963.
4. Vanderbrouke JB. *Observational Research, Randomised Trials, and Two Views of Medical Science*. *PLOS Medicine* 2008;5:339–43.
5. Deming WE. On probability as a basis for action. *Am Stat* 1975;29:146–52.
6. Shewhart WA. *Statistical Method from the Viewpoint of Quality Control*. Washington, DC: The Graduate School, Department of Agriculture, 1939.
7. Fletcher R, Fletcher S, Wagner E. *Clinical Epidemiology: The Essentials*. Boston: Lippincott Williams & Wilkins, 1988.
8. Deming WE. On a classification of the problems of statistical inference. *J Am Stat Assoc* 1942;37:173–85.
9. Deming WE. *Some Theory of Sampling*. New York: John Wiley & Sons, 1950.
10. Moen RM, Nolan TW, Provost LP. *Quality Improvement through Planned Experimentation*. 2nd edn. New York: McGraw-Hill, 1998.
11. Deming WE. *The New Economics for Industry, Government, and Education*. Cambridge, MA: MIT Press, 1993.
12. Baily MA, Bottrell M, Lynn J, et al. *The Ethics of Improving Health Care Quality & Safety: A Hastings Center/AHRQ Project*. Garrison, NY: The Hastings Center, 2004.