

# The meaning of variation to healthcare managers, clinical and health-services researchers, and individual patients

Duncan Neuhauser,<sup>1</sup> Lloyd Provost,<sup>2</sup> Bo Bergman<sup>3</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio, USA

<sup>2</sup>Associates in Process Improvement, Austin, Texas, USA

<sup>3</sup>Centre for Health Improvement, Chalmers University of Technology, Gothenburg, Sweden

## Correspondence to

Charles Elton Blanchard Professor Duncan Neuhauser, Department of Epidemiology and Biostatistics, Medical School, Case Western Reserve University, 10900 Euclid Ave, Cleveland Ohio 44106-4249, USA; dvn@case.edu

Accepted 22 January 2011

## ABSTRACT

Healthcare managers, clinical researchers and individual patients (and their physicians) manage variation differently to achieve different ends. First, managers are primarily concerned with the performance of care processes over time. Their time horizon is relatively short, and the improvements they are concerned with are pragmatic and 'holistic.' Their goal is to create processes that are stable and effective. The analytical techniques of statistical process control effectively reflect these concerns. Second, clinical and health-services researchers are interested in the effectiveness of care and the generalisability of findings. They seek to control variation by their study design methods. Their primary question is: 'Does A cause B, everything else being equal?' Consequently, randomised controlled trials and regression models are the research methods of choice. The focus of this reductionist approach is on the 'average patient' in the group being observed rather than the individual patient working with the individual care provider. Third, individual patients are primarily concerned with the nature and quality of their own care and clinical outcomes. They and their care providers are not primarily seeking to generalise beyond the unique individual. We propose that the gold standard for helping individual patients with chronic conditions should be longitudinal factorial design of trials with individual patients. Understanding how these three groups deal differently with variation can help appreciate these three approaches.

## INTRODUCTION

Health managers, clinical researchers, and individual patients need to understand and manage variation in healthcare processes in different time frames and in different ways. In short, they ask different questions about why and how healthcare processes and outcomes change (table 1). Confusing the needs of these three stakeholders results in misunderstanding.

## HEALTH MANAGERS

Our extensive experience in working with healthcare managers has taught us that their primary goal is to maintain and improve the quality of care processes and outcomes for groups of patients. Ongoing care and its improvement are temporal, so in their situation, learning from variation over time is essential. Data are organised over time to answer the fundamental management question: is care today as good as or better than it was in the past, and how likely is it to be better tomorrow? In answering that question, it becomes crucial to understand the difference between common-cause and special-cause variation (as will be discussed later). Common-cause variation appears as random variation in all measures from healthcare processes.<sup>1</sup> Special-cause variation appears as the effect of causes outside the core processes of the work. Management can reduce this variation by enabling the easy recognition of special-cause variation and by changing healthcare processes—by supporting the use of clinical practice guidelines, for example—but common-cause variation can never be eliminated.

The magnitude of common-cause variation creates the upper and lower control limits in Shewhart control charts.<sup>2–5</sup> Such charts summarise the work of health managers well. Figure 1 shows a Shewhart control chart (p-chart) developed by a quality-improvement team whose aim was to increase compliance with a new care protocol. The clinical records of eligible patients discharged (45–75 patients) were evaluated each week by the team, and records indicating that the complete protocol was followed were identified. The baseline control chart showed a stable process with a centre line (average performance) of 38% compliance. The team



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

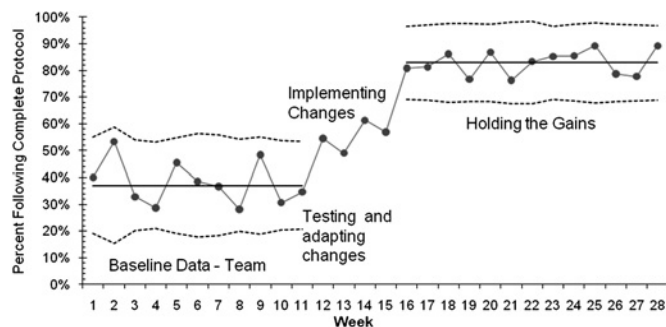
**Table 1** Meaning of variation to managers, researchers and individual patients: questions, methods and time frames

Role	Question	Methods	Time frame	Variation
Health managers	Are we getting better?	Control charts, holistic change	Real time, months' variation	Creating stable processes, learning from special cause
Clinical and health-services researchers	Other things equal, does A cause B?	Randomised controlled trials, regression models; reductionist	Not urgent, years	Eliminate special-cause variation, test for significance, focus on mean values
Individual patient (and provider)	How can I get better?	Longitudinal, factorial designs	Days, weeks, lifelong	Help in understanding the many reasons for variation in health

analysed the aspects of the protocol that were not followed and developed process changes to make it easier to complete these particular tasks. After successfully adapting the changes to the local environment (indicated by weekly points above the upper control limit in the 'Implementing Changes' period), the team formally implemented the changes in each unit. The team continued to monitor the process and eventually developed updated limits for the chart. The updated chart indicated a stable process averaging 83%.

This control chart makes it clear that a stable but inferior process was operating for the first 11 weeks and, by inference, probably before that. The annotated changes (testing, adapting and implementing new processes of care) are linked to designed tests of change which are special (assignable) causes of variation, in this case, to improvement after week 15, after which a new better stable process has taken hold. Note that there is common-cause (random) variation in both the old and improved processes.

After updating the control limits, the chart reveals a new stable process with no special-cause variation, which is to say, no points above or below the control limits (the dotted lines). Note that the change after week 15 cannot easily be explained by chance (random, or common-cause, variation), since the probability of 13 points in a row occurring by chance above the baseline control limit is one divided by 2 to the 13th power. This is the same likelihood that in flipping a coin 13 times, it will come up heads every time. This level of statistical power



**Figure 1** Annotated Shewhart control chart—using protocol.

to exclude randomness as an explanation is not to be found in randomised controlled trials (RCTs). Although there is no hard-and-fast rule about the number of observations over time needed to demonstrate process stability and establish change, we believe a persuasive control chart requires 20–30 or more observations.

The manager's task demonstrates several important characteristics. First is the need to define the key quality characteristics, and choose among them for focused improvement efforts. The choice should be made based on the needs of patients and families. The importance of these quality characteristics to those being served means that speed in learning and improvement is important. Indeed, for the healthcare manager, information for improvement must be as rapid as possible (in real time). Year-old research data are not very helpful here; just-in-time performance data in the hands of the decision-makers provide a potent opportunity for rapid improvement.<sup>6</sup>

Second, managerial change is holistic; that is, every element of an intervention that might help to improve and can be done is put to use, sometimes incrementally, but simultaneously if need be. Healthcare managers are actively working to promote measurement of process and clinical outcomes, take problems in organisational performance seriously, consider the root causes of those problems, encourage the formation of problem solving clinical micro-system teams and promote the use of multiple, evolving Plan–Do–Study–Act (PDSA) tests of change.

This kind of improvement reasoning can be applied to a wide range of care processes, large and small. For example, good surgery is the appropriate combination of hundreds of individual tasks, many of which could be improved in small ways. Aggregating these many smaller changes may result in important, observable improvement over time. The protocol-driven, randomised trial research approach is a powerful tool for establishing efficacy but has limitations for evaluating and improving such complex processes as surgery, which are continually and purposefully changing over time. The realities of clinical improvement call for a move from after-the-fact

quality inspection to building quality measures into medical information systems, thereby creating real-time quality data for providers to act upon. Caring for populations of similar patients in similar ways (economies of scale) can be of particular value, because the resulting large numbers and process stability can help rapidly demonstrate variation in care processes<sup>7</sup>; very tight control limits (minimal common-cause variation) allow special-cause variation to be detected more quickly.

### CLINICAL AND HEALTH-SERVICES RESEARCHERS

While quality-management thinking tends towards the use of data plotted over time in control-chart format, clinical researchers think in terms of true experimental methods, such as RCTs. Health-services researchers, in contrast, think in terms of regression analysis as their principal tool for discovering explainable variation in processes and outcomes of care. The data that both communities of researchers use are generally collected during fixed periods of time, or combined across time periods; neither is usually concerned with the analysis of data over time.

Take, for example, the question of whether age and sex are associated with the ability to undertake early ambulation after hip surgery. Clinical researchers try to control for such variables through the use of entry criteria into a trial, and random assignment of patients to experimental or control group. The usual health-services research approach would be to use a regression model to predict the outcome (early ambulation), over hundreds of patients using age and sex as independent variables. Such research could show that age and sex predict outcomes and are statistically significant, and that perhaps 10% of the variance is explained by these two independent variables. In contrast, quality-improvement thinking is likely to conclude that 90% of the variance is unexplained and could be common-cause variation. The health-services researcher is therefore likely to conclude that if we measured more variables, we could explain more of this variance, while improvement scientists are more likely to conclude that this unexplained variance is a reflection of common-cause variation in a good process that is under control.

The entry criteria into RCTs are carefully defined, which makes it a challenge to generalise the results beyond the kinds of patients included in such studies. Restricted patient entry criteria are imposed to reduce variation in outcomes unrelated to the experimental intervention. RCTs focus on the difference between point estimates of outcomes for entire groups (control and experimental), using statistical tests of significance to show that differences between the two arms of a trial are not likely to be due to chance.

### INDIVIDUAL PATIENTS AND THEIR HEALTHCARE PROVIDERS

The question an individual patient asks is different from those asked by manager and researcher, namely 'How can I get better?' The answer is unique to each patient; the question does not focus on generalising results beyond this person. At the same time, the question the patient's physician is asking is whether the group results from the best clinical trials will apply in this patient's case. This question calls for a different inferential approach.<sup>8-10</sup> The cost of projecting general findings to individual patients could be substantial, as described below.

Consider the implications of a drug trial in which 100 patients taking a new drug and 100 patients taking a placebo are reported as successful because 25 drug takers improved compared with 10 controls. This difference is shown as not likely to be due to chance. (The drug company undertakes a multimillion dollar advertising campaign to promote this breakthrough.) However, on closer examination, the meaning of these results for individual patients is not so clear. To begin with, 75 of the patients who took the drug did not benefit. And among those 25 who benefited, some, perhaps 15, responded extremely well, while the size of the benefit in the other 10 was much smaller. To have only the 15 'maximum responders' take this drug instead of all 100 could save the healthcare system 85% of the drug's costs (as well as reduce the chance of unnecessary adverse drug effects); those 'savings' would, of course, also reduce the drug company's sales proportionally. These considerations make it clear that looking at more than group results could potentially make an enormous difference in the value of research studies, particularly from the point of view of individual patients and their providers.

In light of the above concerns, we propose that the longitudinal factorial study design should be the gold standard of evidence for efficacy, particularly for assessing whether interventions whose efficacy has been established through controlled trials are effective in individual patients for whom they might be appropriate (box 1). Take the case of a patient with hypertension who measures her blood pressure at least twice every day and plots these numbers on a run chart. Through this informal observation, she has learnt about several factors that result in the variation in her blood pressure readings: time of day, the three different hypertension medicines she takes (not always regularly), her stress level, eating salty French fries, exercise, meditation (and, in her case, saying the rosary), and whether she slept well the night before. Some of these factors she can control; some are out of her control.

**Box 1** Longitudinal factorial design of experiments for individual patients

The six individual components of this approach are not new, but in combination they are new<sup>8 9</sup>

1. One patient with a chronic health condition; sometimes referred to as an 'N-of-1 trial.'
2. Care processes and health status are measured over time. These could include daily measures over 20 or more days, with the patient day as the unit of analysis.
3. Whenever possible, data are numerical rather than simple clinical observation and classification.
4. The patient is directly involved in making therapeutic changes and collecting data.
5. Two or more inputs (factors) are experimentally and concurrently changed in a predetermined fashion.
6. Therapeutic inputs are added or deleted in a predetermined, systematic way. For example: on day 1, drug A is taken; on day 2, drug B; on day 3, drug A and B; day 4, neither. For the next 4 days, this sequence could be randomly reordered.

Since she is accustomed to monitoring her blood pressure over time, she is in an excellent position to carry out an experiment that would help her optimise the effects of these various influences on her hypertension. Working with her primary care provider, she could, for example, set up a table of randomly chosen dates to make each of several of these changes each day, thereby creating a systematically predetermined mix of these controllable factors over time. This factorial design allows her to measure the effects of individual inputs on her blood pressure, and even interactions among them. After an appropriate number of days (perhaps 30 days, depending on the trade-off between urgency and statistical power), she might conclude that one of her three medications has no effect on her hypertension, and she can stop using it. She might also find that the combination of exercise and consistently low salt intake is as effective as either of the other two drugs. Her answers could well be unique to her. Planned experimental interventions involving single patients are known as 'N-of-1' trials, and hundreds have been reported.<sup>10</sup> Although longitudinal factorial design of experiments has long been used in quality engineering, as of 2005 there appears to have been only one published example of its use for an individual patient.<sup>8 9</sup> This method of investigation could potentially become widely used in the future to establish the efficacy of specific drugs for individual patients,<sup>11</sup> and perhaps even required, particularly for very expensive drug therapies for chronic conditions. Such individual trial results could be combined to obtain generalised knowledge.

This method can be used to show (1) the independent effect of each input on the outcome, (2) the interaction

effect between the inputs (perhaps neither drug A or B is effective on its own, but in combination they work well), (3) the effect of different drug dosages and (4) the lag time between treatment and outcome. This approach will not be practical if the outcome of interest occurs years later. This method will be more practical with patient access to their medical record where they could monitor all five of Bergman's core health processes.<sup>12</sup>

### UNDERSTANDING VARIATION IS ONE OF THE CORNERSTONES OF THE SCIENCE OF IMPROVEMENT

This broad understanding of variation, which is based on the work of Walter Shewart in the 1920s, goes well beyond such simple issues as making an intended departure from a guideline or recognising a meaningful change in the outcome of care. It encompasses more than good or bad variation (meeting a target). It is concerned with more than the variation found by researchers in random samples from large populations.

Everything we observe or measure varies. Some variation in healthcare is desirable, even essential, since each patient is different and should be cared for uniquely. New and better treatments, and improvements in care processes result in beneficial variation. Special-cause variation should lead to learning. The 'Plan-Do-Study' portion of the Shewart PDSA cycle can promote valuable change.

The 'act' step in the PDSA cycle represents the arrival of stability after a successful improvement has been made. Reducing unintended, and particularly harmful, variation is therefore a key improvement strategy. The more variation is controlled, the easier it is to detect changes that are not explained by chance. Stated differently, narrow limits on a Shewart control chart make it easier and quicker to detect, and therefore respond to, special-cause variation.

The goal of statistical thinking in quality improvement is to make the available statistical tools as simple and useful as possible in meeting the primary goal, which is not mathematical correctness, but improvement in both the processes and outcomes of care. It is not fruitful to ask whether statistical process control, RCTs, regression equations or longitudinal factorial design of experiments is best in some absolute sense. Each is appropriate for answering different questions.

### Forces driving this new way of thinking

The idea of reducing unwanted variation in healthcare represents a major shift in thinking, and it will take time to be accepted. Forces for this change include the computerisation of medical records leading to public reporting of care and outcome comparisons between providers and around the world. This in turn will

promote pay for performance, and preferred provider contracting based on guideline use and good outcomes. This way of thinking about variation could spread across all five core systems of health,<sup>12</sup> including self-care and processes of healthy living.

**Competing interests** None.

**Provenance and peer review** Not commissioned; externally peer reviewed.

### REFERENCES

1. Lighter DE. *Advanced Performance Improvement in Health Care*. Sudbury, MA: Jones and Bartlet, 2011, Chapter 5.
2. Bergman B, Klefjso B. *Quality from Customer Needs to Customer Satisfaction*. 3rd edn. Lund, Sweden: Studentlitteratur, 2010.
3. Berwick D. Controlling variation in health care: a consultation with Walter Shewhart. *Med Care* 1991;29:1212–25.
4. Moen RM, Nolan TW, Provost LP. *Quality Improvement through Planned Experimentation*. 2nd edn. New York: McGraw-Hill, Inc 1998.
5. Nolan TW, Provost LP. *Understanding Variation, Quality Progress*. ASQC, 1990.
6. Lifvergren S, Gremyer I, Hellstrom A, *et al*. Lessons from Sweden's First Large Scale Implementation of Six Sigma in Health Care. *Oper Manage Res* 2010;3:117–28.
7. Diaz M, Neuhauser D. Pasteur and parachutes: when statistical process control is better than a randomized clinical trial. *Qual Saf Health Care* 2005;14:140–3.
8. Thor J. *Getting Going on Getting Better: How is Systematic Quality Improvement Established in a Health Care Organization—Implications for Change Management and Theory*. Doctoral dissertation. Stockholm: Medical Management Centre, Karolinska Institute, 2007 and summarized in Quality and Safety in Health Care.
9. Olsson J, Terris D, Elg M, *et al*. The one person randomized controlled trial. *Qual Manag Health Care* 2005;14:206–16.
10. Larsen EB. N-of-1 Trials: a new future? *J Gen Intern Med* 2010;25:891–2.
11. Hebert C, Neuhauser D. Improving hypertension care with patient generated run charts. *Qual Manag Health Care* 2004;13:174–7.
12. Bergman B, Neuhauser D, Provost L. Five main processes in healthcare—a citizen perspective. *BMJ Qual Saf* 2011;20(S1): i41–i42.