

## Principles of quality improvement

# Data ‘sanity’: statistics and reality

Davis Balestracci MS

Harmony Consulting, Portland, USA

### ABSTRACT

A quality improvement context invalidates many assumptions perpetuated by traditional research-oriented statistical thinking, thereby rendering commonly used analyses inappropriate. The science of quality improvement is radically different from clinical research. In addition, the stigma resulting from past (often compulsory) statistics courses often presents a formidable cultural barrier to the much-needed simple, efficient data collection and

analysis methods which are key to improving existing healthcare processes in real time. Proper implementation of statistical thinking has implications for and far beyond clinical outcomes.

**Keywords:** data process, ‘enumerative’/‘analytic’ statistics, prediction, process-oriented thinking, random sampling, variation

## Introduction

Required ‘Statistics from Hell 101’ courses are virtually *worthless* ... and perpetuate the myth that statistics can be used to ‘massage’ data and prove anything. Common statistical techniques used on organisational data such as variance tables, trends analysis, rankings, stretch goals, and tougher standards, often used inappropriately, can actually *sabotage* improvement!

*Whether or not you understand statistics, you are already using statistics!* The key skill needed is the ability to respond to variation *appropriately* so as to *ask better questions* – mathematical skills aren’t necessarily the focus!

## Data ‘sanity’: statistical thinking applied to everyday data

Most people generally do not perceive that they need statistics; their need is first and foremost to solve problems.

Given the current rapid pace of change in the healthcare environment along with the ‘benchmarking’, ‘re-engineering’, ‘total customer satisfaction’, and, most recently, ‘Six Sigma’ and ‘Lean’ crazes, there seems to be a new and increasing tendency for performance

goals to be imposed from external sources, making improvement efforts flounder when:

- results are presented in aggregated row and column formats complete with variances and rankings
- perceived trends are acted upon to reward and punish
- labels such as ‘above average’ and ‘below average’ get attached to individuals or institutions
- stakeholders are ‘outraged’ by certain results and impose even ‘tougher’ standards.

These are very well-meaning strategies that are simple, obvious ... and wrong! They will mislead analysis and interpretation ... and insidiously cloud decisions every day in virtually every work environment.

The realities are:

- taking action to improve a situation is tantamount to using statistics
- ‘traditional’ statistics have severely limited value in real-world settings
- understanding of variation is more important than using specific techniques
- statistical thinking gives a knowledge base from which to ask the right questions
- unforeseen problems are caused by the exclusive use of arbitrary numerical goals, ‘stretch’ goals, and ‘tougher’ standards for driving improvement

- using heavily aggregated tables of numbers, variances from budgets, or bar graph formats as tools for taking meaningful management action are many times futile and inappropriate
- there is poor awareness of the true meaning of 'trend', 'above average' and 'below average'.

## A key concept: process-oriented thinking

The statistics needed for quality improvement are based in the context of process. What is a process? All work is a process! Processes are sequences of tasks aimed at accomplishing a particular outcome by manipulating inputs to produce a particular type of output. Everyone involved in the process has a role of supplier, processor or customer. A group of related processes is called a system.

*Process-oriented thinking* is built on the following premises:

- understanding that:
  - *all* work is accomplished through a series of one or more processes, each of which is potentially measurable
  - all processes exhibit *variation*, which inhibits their predictability
  - if a process does not 'go right' it exhibits undesirable variation
  - processes 'speak' to us through data
  - processes are *perfectly designed* to get the results they are already getting ... even if they're getting results they 'shouldn't' be getting!
- process inputs falling into the six general categories of 'people', 'methods', 'machines', 'materials',

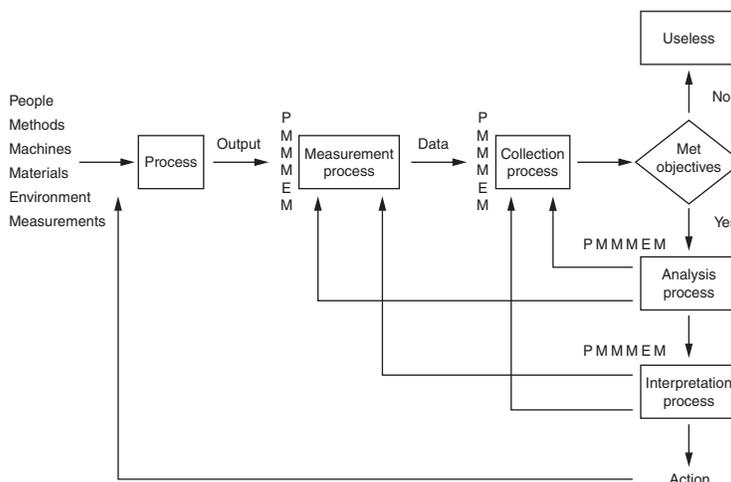
'measurements' (data) and 'environment', each of which is a potential source of variation

- reducing inappropriate and unintended variation by:
  - eliminating work procedures that do not add value (i.e. only add cost with no payback to customers)
  - ensuring that all workers are performing at the best inherent level of the process's capability with their given inputs
  - reacting appropriately to variation because there are two types – treating one as the other will actually make things worse
- improving quality = improving processes (more consistent prediction).

Aside from the new perspective of looking at your jobs and workplaces as processes and systems, process-oriented thinking must also be applied to a quality professional's *data collection process*.

## How a quality professional 'adds value': recognising the use of data as a process

The use of data is really made up of four processes – measurement, collection, analysis and interpretation – each having 'people', 'methods', 'machines', 'materials', 'measurements' (numbers) and 'environment' as inputs (see Figure 1). *Any one of these six inputs can be a source of variation for any one of these four processes* – they lurk to contaminate your data process and mislead you as to what is going on in the actual system you are trying to improve!



**Figure 1** People, methods, machines, materials, environment and measurements inputs can be a source of variation for any one of the measurement, collection, analysis or interpretation processes!

So, any process produces outputs that are potentially measurable. If one chooses, one can obtain a number (a piece of data) characterising the situation through a process of *measurement*, called an *operational definition*. If the objectives are not understood or people have varying perceptions of what is being measured, the six sources of variation will compromise the quality of this measurement process.

For example: (1) How many 'beds' does your hospital have? (2) How many 'patient deaths' occurred last year? (3) Define 'stopped smoking'. W Edwards Deming was fond of saying, 'There is no true value of anything'. Crude measures of the right things are better than precise measures of the wrong things – as long as it's 'consistently inconsistent' and defined in a way so that all will get the same number, you will benefit from the elegant simplicity of the statistical techniques inherent in quality improvement.

These individual measurements must then be accumulated into an appropriate dataset, so they next pass to a *collection* process. If the objectives are clear, the designed collection process should be relatively well defined *because the analysis is known ahead of time* – the appropriateness of an analysis depends on how the data were collected. If the objectives are not clear, the six sources of variation will once again act to compromise the process. (Actually, from the author's experience, it is virtually guaranteed that the six sources will compromise the collection process anyway!)

If the objectives are passive and reactive, eventually someone will extract the data and use a computer to 'get the stats'. This, of course, is an *analysis* process (albeit not necessarily a good one) that also has the six sources of inputs as potential sources of variation. Or, maybe more commonly, someone extracts the data and hands out tables of raw data presented as computer-generated summary analyses at a meeting. This becomes the analysis process, which is affected by the variation in perceptions and abilities of people at the meeting.

Ultimately, as you now are starting to realise, it all boils down to interpreting the variation in the measurements. So, the *interpretation* process (with the same six sources of inputs) results in an *action* that is then fed back in to the original process.

Now, think back for a minute to the many meetings you attend. How do unclear objectives, inappropriate or misunderstood data definitions, unclear or inappropriate data collections, passive statistical 'analyses', and shoot-from-the-hip interpretations of variation influence the agendas and action? In fact, *how many times are people merely reacting to the variation in these elements of the data process* – and making decisions that have **nothing** to do with the process being studied?

Another danger inherent in this data process is that data not collected *specifically* for the current objective can generally be 'tortured' to 'confess' to someone else's hidden agenda!<sup>1</sup>

Given the nature of process-oriented thinking, one of the biggest changes in thinking will be to realise the benefit of studying a process by collecting more frequent samples over time, which will cause the need to redefine many current pristine organisational operational definitions to become 'good enough'.

## Research versus improvement

So, if the data process itself is flawed, many hours are spent 'spinning wheels' due to the contamination from the 'human' variation factors inherent in the aforementioned processes – people make decisions and react to their perceptions of the *data* process rather than the process allegedly being improved!

Many doctors will argue that the process-oriented approach is invalid because it doesn't follow 'established' procedures of clinical research. However, let's look at *research as a process*.

In research, all input variations are tightly controlled such that observed differences in the 'control' and 'treatment' groups can be attributed to the 'methods' input and no other. The protocol is also excruciatingly detailed as to how measurements are defined, collected, and analysed so as to reduce variation in the data process – and it is all defined before one patient is randomised. Controlling this variation is expensive, which is why research is expensive.

Hence, research statistics are actually a very specialised subset of process-oriented statistics! It makes the assumption and has the luxury of 'ignoring' the everyday factors lurking to compromise results in busy uncontrolled naturalistic practice environments.

However, after a significant result is published, do the researchers have any control over how clinicians use it? The 'rigour' is gone, and human variation in interpretation and use of the protocol virtually guarantees that they won't achieve the same results as reported. This variation, usually either inappropriate or unintended, could be present in any or all of the six inputs ... of five processes (actual and data measurement, collection, analysis, and interpretation)!

So, in understanding any variation between the research results and actual results, it becomes necessary *to expose the variation* between individual use and the research use of the protocol and reduce any inappropriate and unintended variation. As will be discussed in the next sections, traditional statistical methods are, for the most part, invalidated.

Do not underestimate the factors lurking in the data process that will contaminate and invalidate statistical analyses. *Objectives* are crucial for properly *defining* a situation and determining how to *collect* the data for the *appropriate analysis*. Statistical theory can interpret the *variation* exposed by the analysis to take *appropriate action*.

## Further clarification: 'enumerative' statistics versus 'analytic' statistics

(I am indebted to the work of David Kerridge for much of the following explanation as well as those in the subsequent sections 'Prediction' and 'Unknown and unknowable'. Kerridge D. *Statistics and Reality*. Unpublished manuscript.)

There are actually three kinds of statistics, and they can be summarised as follows:

- *descriptive*: 'What can I say about this *individual* patient?'
- *enumerative*: 'What can I say about this *specific group* of patients?'
- *analytic*: 'What can I say about the *process* that produced the result in this group of patients?'

An *enumerative study* always focuses on the *actual state* of something *at one point in the past* – no more, no less. For example, one can literally summarise the results of all the participants in any clinical trial once it is completed.

An analytic study usually focuses on predicting the results of action in the future – in circumstances we cannot fully know. It is this predictive way of thinking that is fundamental to quality improvement.

Both kinds of statistics count or measure samples. However, suppose it is desired to know which of two antibiotics is better in treating a certain disease. It is impossible to take a random sample of all the people who will be treated in the future: it isn't even known who specifically will get this disease in the future!

This can be described as sampling from an *imaginary population*. The practical difference is that we must not rely on what happens from the results of any one experiment: we must repeat the experiment under as many different circumstances as we can to establish an increasing degree of belief in the result. This is in very strong contrast with what is normally taught in most statistics textbooks, which describe the problem as one of 'accepting' or 'rejecting' hypotheses.

Walter Shewhart stated the difference by means of an example:<sup>2</sup>

'You go to your tailor for a suit of clothes and the first thing that he does is to make some measurements; you go

to your physician because you are ill and the first thing he does is to make some measurements. The objects of making measurements in these two cases are different. They typify the two general objects of making measurements. They are:

- (a) to obtain quantitative information
- (b) to obtain a causal explanation of observed phenomena.'

## Prediction

The distinction between enumerative and analytic studies means we must look for repeatability over many different populations consistently over time. Analytic thinking relates to sampling from a *process*, rather than a well-defined, finite *population*. Furthermore, most mathematical statisticians state statistical problems in terms of repeated sampling from the *same population* under circumstances where nothing changes over time! This leads to a very simple mathematical theory, but does not relate to the real needs of the statistical user. Especially in medicine, one cannot take repeated samples from the exactly the same population, except in rare cases.

Getting back to comparing two antibiotics in the treatment of some infection, suppose a conclusion is made that one did better in tests. How does that help?

Suppose that all testing was done in one hospital in New York in 2003; however, someone may want to use the same antibiotic in Africa in 2006. It is quite possible that the best antibiotic for New York is not the same as the best in a refugee camp in Zaire. In New York the strains of bacteria may be different: and the problems of transport and storage really are different. If the antibiotic is freshly made and stored in efficient refrigerators, it may be excellent. It may not work at all if transported to a camp with poor storage facilities.

And even if the same antibiotic works in both places, how long will it go on working? This will depend on how carefully it is used and how quickly resistant strains of bacteria build up. The effectiveness of a drug may also depend on the age of the patient, or previous treatment, or the stage of the disease. Ideally it is desirable to have one treatment that works well in all foreseeable circumstances, but this may not be possible.

## Unknown and unknowable

There are usually no difficulties carrying out the objectives of an enumerative study, which often involves estimation, other than how to choose the sample; however, in quality improvement, an analytic process, a number of subsequent practical problems still remain.

Random sampling is often used in analytic studies, but this is not the same as sampling in an enumerative study. For example, consider a chosen group of patients with hypertension included in a randomised controlled study, who attend a particular clinic. Either a random method or some complicated method involving random numbers is used to determine who is to get which treatment. But the resulting sample is *not necessarily a random sample of the patients who will be treated in the future at that same clinic.*

Still less are they a random sample of the patients who will be treated in any other clinic. In fact the patients who will be treated in the future will depend on choices that you and others have not yet made! And those choices will depend on the results of the study currently being done, and on studies by other people that may be carried out in the future.

So with an analytic study, there are two distinct sources of uncertainty. The first is similar to an enumerative study, that due to sampling.

The second is due to the fact that one is predicting what will happen at some time in the future – to some group that is different from the original sample. This uncertainty is ‘*unknown and unknowable*’. It is rarely known how any produced results will be used, so all one can do is to warn the potential user of the range of uncertainties that might affect different actions.

This is rarely done. Furthermore, how does one even express it? But the uncertainties of this kind will in most circumstances be an order of magnitude greater than the uncertainty due merely to sampling – making it very dangerous to pretend to be more certain than warranted. Such false certainty many times leads to wrong choices, but the result, in most statistics courses, has been a theory in which the unmeasured uncertainty has just been ignored.

## ‘Statistics’ in a quality improvement perspective

So, it can be seen that statistics is not merely the science of analysing data, but the art and science of collecting and analysing data. Given any improvement situation (including daily work), one must be able to:

- 1 choose and define the problem in a *process and systems* context
- 2 design and manage a series of *simple, efficient data collections* to expose undesirable variation

- 3 use comprehensible methods presentable and understandable across all layers of the organisation, virtually all *graphical* avoiding raw data or bar graphs (with the specific exception of a Pareto analysis), to expose further the inappropriate and unintended variation
- 4 numerically *assess* the current state of an undesirable situation, further expose inappropriate and unintended variation, assess the effects of interventions
- 5 *hold the gains* of any improvements made, generally requiring a much simpler data collection.

## Summary

As quality professionals, it is important to realise that data analysis goes far beyond the routine statistical ‘crunching’ of numbers. The greatest contribution to an organisation is getting people to understand and use a *process-oriented* context in analysing situations as well as principles of good, *simple, efficient* data collection, analysis, and display. This cannot help but enhance the healthcare quality professional’s credibility. It will also help gain the confidence and cooperation of organisations during stressful transitions and external assessments.

*Whether or not people understand statistics, they are already using statistics ...* and with the best of intentions. It is therefore vital to put a stop to many of the current well-meaning but ultimately damaging *ad hoc* uses of statistics.

## REFERENCES

- 1 Mills JL. Data torturing. *New England Journal of Medicine* 1993;329:1196–9.
- 2 Shewhart W. *The Economic Control of Manufactured Product*. Milwaukee, WI: American Society for Quality, 1980.

## ADDRESS FOR CORRESPONDENCE

Davis Balestracci, Harmony Consulting, 94 Ashley Ln, Portland, ME 04103-2789, USA. Tel: +207 899 0962; email: [davis@dbharmony.com](mailto:davis@dbharmony.com); website: [www.dbharmony.com](http://www.dbharmony.com)

Received 18 January 2006

Accepted 6 February 2006