Controlling Variation in Health Care: A Consultation from Walter Shewhart
Author(s): Donald M. Berwick
Source: *Medical Care*, Vol. 29, No. 12 (Dec., 1991), pp. 1212-1225
Published by: Lippincott Williams & Wilkins
Stable URL: http://www.jstor.org/stable/3765991
Accessed: 24-12-2015 14:44 UTC

MEDICAL CARE
December 1991, Vol. 29, No. 12

Controlling Variation in Health Care:

A Consultation from Walter Shewhart

DONALD M. BERWICK, MD, MPP

The control of unintended variation is an objective central to modern industrial quality management methods, based largely on the theoretical work of Walter A. Shewhart. As industrial quality management techniques find their place in health care, professionals may feel threatened by the effort to reduce variation. Understanding may reduce this fear. Variation of the types addressed in quality control efforts erodes quality and reliability, and adds unnecessarily to costs. Such undesirable variation derives, for example, from misinterpretation of random noise in clinical data, from unreliability in the performance of clinical and support systems intended to support care, from habitual differences in practice style that are not grounded in knowledge or reason, and from the failure to integrate care across the boundaries of components of the health care system. Quality management efforts can successfully reduce each of these forms of variation without insult to the professional autonomy, dignity, or purpose of health care professionals. Professionals need to embrace the scientific control of variation in the service of their patients and themselves. Key words: quality assurance; quality control; quality improvement; variation; protocols. (Med Care 1991; 29:1212–1225)

The Lines of Cause

Kim, aged 3 years, lies asleep, waiting for a miracle. Outside her room, the nurses on the night shift pad softly through the half-lighted corridors, stopping to count breaths, take pulses, or check the intravenous pumps. In the morning, Kim will have her heart fixed. She will be medicated and wheeled into the operating suite. Machines will take on the functions of her body: breathing and circulating blood. The surgeons will place a small patch over a hole within her heart, closing off a shunt between her ventricles that would, if left open, slowly kill her.

Kim will be fine if the decision to operate on her was correct; if the surgeon is competent; if that competent surgeon happens to be trained to deal with the particular anatomic wrinkle that is hidden inside Kim's heart; if the blood bank cross-matched her blood accurately and delivered it to the right place; if the blood gas analysis machine works properly and on time; if the suture does not snap; if the plastic tubing of the heart-lung machine does not suddenly spring loose; if the recovery room nurses know that she is allergic to penicillin; if the "oxygen" and "nitrogen" lines in the anesthesia machine have not been reversed by mistake; if the sterilizer temperature gauge is calibrated so that the instruments are in fact

From the Department of Pediatrics, Harvard Medical School, Boston, Massachusetts.

This paper was first presented at the 75th Anniversary Symposium, Henry Ford Health Systems, Detroit, Michigan, November 7, 1990.

Address correspondence to: Donald M. Berwick, MD, MPP, 131 Lake Avenue, Newton, MA 02159.

sterile; if the pharmacy does not mix up two labels; and if when the surgeon says urgently, "Clamp, right now," there is a clamp on the tray.

If all goes well, if ten thousand "ifs" go well, then Kim may sing her grandchildren to sleep some day. If not, she will be dead by noon tomorrow.

If Kim were an astronaut, strapped into her seat at the top of some throbbing rocket, the crowd assembled would hold their breath in the morning Florida sun. "How can it possibly work?" they would whisper. "How many parts are there in that machine? A million? What if one fails? My *toaster* fails. Please let it all work right." The machine would bellow smoke, the gantry fall away, and slowly the monster would rise, Kim on top.

If it worked, they would cheer. "A miracle," they would shout, in awe that the millions of tiny lines of effort, the millions of tiny lines of cause and effect, from job shops in Ohio and laboratories in Pasadena, crisscrossing through time and space, could converge so magnificently in a massive, gleaming rocket launched *exactly right*. Perfect.

If it failed, they would cry. So would the rocket's makers, who had done their very best. No one wanted it to end this way. Poor Kim. What was the trouble? What went wrong? Why?

The lines of cause will converge around Kim in the morning as she wheels toward the operating room. Thousands upon thousands of elements weaving a basket to hold her safely, all hope. No crowd holds its breath tonight; but wouldn't they if they knew?

The Illusion of Control

As I do once a year, I had the privilege several months ago to serve as an attending physician at a superb tertiary children's hospital. The experience of trying to teach in that setting is always humbling. I feel embedded in some immense, whirring machine, spinning around me no less than it spins around Kim. I am allegedly in some control, control that is indicated by such terms as "supervision," "attending rounds," and "doctor's orders."

But, in truth, these terms are euphemisms. I ratify, perhaps, or I assent, but "control" is too strong a term for what I do. My questions, my requests, and my instructions may result in some slight adjustments of direction, but the juggernaut rolls on for the most part quite oblivious of me. Kim and I are both passengers. Who is steering? I don't know. Habit, maybe? Convention? Rumor? Perhaps higher, hidden authorities?

I stop to ask the senior resident about the sudden prevalence of pulse oximetry in the management of asthmatics. A half-dozen pulse oximeters are in use this very morning. I do not recall this from my own training; nor do I understand its logic.

"Does pulse oximetry really make a difference?" I ask him.

"We use it now," he answers. He does not specify the antecedent for the pronoun, "We."

It took me a month, but now I know who really controls events in the modern hospital. It is "we," the pronoun with no antecedent. "We," as in, "We believe that you need a biopsy, Mr. Fowler," or "We use aminophylline drips," or "We don't think you are ready to go home yet," or "We changed antibiotics yesterday because she spiked a fever." I am reassured."We" are in charge or, perhaps more grammatically, "We" *is* in charge. The order form, which I sign, says I am in control. Unluckily, I discover, I am not. Luckily, I discover, "We" is.

"We" will make Kim safe. She will live because "we" plan it.

Nonsense. You know, as I do, that no such plan exists. The "we" without the antecedent is not a conscious, organized, logical, scientifically driven being, individual or group. You know as well as I do that, on the whole, it is a lumbering, unconscious presence, a gigantic, inchoate collective, a system of causes that no one really knows, and that to

attribute "planfulness" to that system is the same as saying that the Colorado River dug the Grand Canyon because it wanted to.

I mean to blame no one in saying this. It is hard to find in any modern organization a more benign, dedicated, intelligent, and generous collection of people than those in an American hospital. It is a privilege to work with them, and it is primarily through them that the American health care system will, I am confident, even yet be rescued.

They are not, however, in control of their own work. Like me, they push at the sides of the work, nudging it toward the perfection they really desire, and, like me, they feel it move only ever so slightly in response to their strenuous efforts. They want it to be better; but they do not know how to make it so.

Total Quality Management

Taming Processes

Into this landscape of frustration there has lately arrived a newcomer to health care, a collection of managerial disciplines developed and widely adopted in other industries and able in those settings to yield products and services of unprecedented quality, value, and reliability.[1-5] The methods go under many different names; one of them is "total quality management." No matter what the approach is called, it consists, at a minimum, of three essential elements: 1) efforts to know the customer ever more deeply and to link that knowledge ever more closely to the day-to-day activities of the organization; 2) efforts to mold the culture of the organization, largely through the deeds of leaders, to foster pride, joy, collaboration, and scientific thinking; and, finally, 3) efforts to continuously increase knowledge of and control over variation in the processes of work through the widespread use of the scientific methods of collection, analysis, and action upon data.

When all these three efforts are developed in synchrony in an organization, continuous improvement flourishes, quality grows, customers are better served, workers feel more pride, and "we" means something. Ask in such an organization why something is done a certain way, and you get answers, not pronouns. The change is so profound that it is sometimes called a "transformation."

The object of total quality management is to give identity to the pronoun, "we." It is to tame the beast of unintended variation. It is to place under benign and well-intended control the full force of production that lies within the organization so that each productive step, each investment of resource, each call upon an individual human worker serves the purpose of the place.

Deeply embedded in the transformation implied in total quality management, however, lies an apparent paradox. On the one hand, proponents of total quality management say that organizations must liberate the talent, imagination, and initiative of those who work in complex systems. Those who serve Kim know how to serve her even better if they are just given the chance. Quality management sounds like it involves the loosening of control. "Empowerment," some call it.

However, few words appear more commonly in the vocabulary of quality management than the word, "control." The word implies predictability, reliability, reduction of variation. It implies measurement, standardization, and regularity. Where, then, is initiative? Where is creativity? Where is empowerment? How can one create an organization that mobilize the inventiveness of everyone in it and at the same time keeps work in a state of statistical control?

That paradox has a short-circuit connection to fear. Physicians, buffeted by regulation, fear handcuffs that will deny them sensible courses of treatment. Hospitals fear mindless inspections to see if they are in line with others. The word "control" charges the discussion, and reason flees.

But here is the fact: Kim's safety—her life, perhaps—depends exactly on the combina-

tion of freedom and control that at first seems so oxymoronic. In fact, there is no paradox; to free ourselves from senseless contention and to get about the real job of improving, we in health care must come to understand fully why that is so. Afraid of control, it turns out, we will remain not free but helpless.

Few concepts give rise to as much fear in medical organizations as "control of variation," but few components in the technology of managing quality are more important. The effort to control variation must occupy a central place in the agenda of total quality management in health care. But, we will have trouble pursuing that agenda unless we pull the fangs of the terminology.

To understand the deep relationship between the improvement of quality and the control of variation, one must uncover the statistical roots of the science of quality management. There is no better place to begin than with the very person who first set out the technical theory that forever after linked quality and control: Walter A. Shewhart.

The Lessons of Walter A. Shewhart:

A Typology of Cause

Shewhart was trained as a physicist, and he spent most of his professional career at Bell telephone laboratories, from 1925 through 1956, where he assembled a group of engineers and statisticians who together crafted the scientific foundations of modern quality control.[6-8] They began by trying to make better telephones; they ended by teaching the world of industry how to do better work.

Walter Shewhart was a student of, above all, causes. He believed that results in complex systems did not just happen but were the consequences of lawful relationships; maybe it was because he was a physicist that he chose to interpret production that way. He believed that, properly analyzed, experience in real causal systems could teach a great deal about those systems, and he devoted much of his professional career to developing methods through which the study of variation in measured results could teach the observer about the causal systems that led to those results. If he had been a physician, he would have been called an applied epidemiologist, or a clinical researcher—and a master at it.

The causal systems that intrigued Shewhart he called "systems of chance cause,"[6] but he used the word "chance" in a most unusual way: to Shewhart, "chance causes" meant, exactly, "unknown causes." It dawned on him that real, unknown causes, were of two distinct types: as he put it, not all systems of chance causes are alike. In particular, some such causal systems produced effects that obeyed understandable mathematical laws. That was fortunate, since, because they obeyed mathematical laws, they permitted one to make predictions based on experience. He called these "constant systems of chance causes," and they are the same as Deming later called "common causes,"[2] and Juran called "random causes."[1]

Constant Systems of Chance Cause (Common Causes of Variation)

Anyone who has studied statistical mechanics is familiar with the connection between randomness and predictability. It is not possible to predict the location or velocity of any specific molecule in a gas; its journey is random. But it is quite possible to predict with great precision the behavior of a system of many such random molecules because their collective effect can be described with a mathematical law. It is a wonderful fact, and not entirely intuitively obvious, that certain systems of unknown causes produce actual, real-world phenomena that closely follow theoretical statistical distributions, e.g., binomials, Poissons, normal, and so on.

Shewhart saw that the same was true in complex production processes. Even though one remained profoundly ignorant of the causal relationships in a real production pro-

cess, the results of a process could be predictable in a statistical sense if that system of causes was constant.

Assignable Causes (Special Causes of Variation)

That led Shewhart to another postulate: namely, that some unknown causes (he called them "assignable") made prediction impossible. These were causes that were not constant, and because they came and went capriciously, they produced results that failed to follow mathematical laws and, therefore, results that could not be predicted based on such laws. Processes with these erratic, nonsystematic causes were unstable, and their performance could not be well characterized with statistics that implied mathematical regularity, such as averages and standard deviations. The processes, to use the expression that became conventional, were not "in statistical control."

It is easier to understand Shewhart's distinctions among types of causal systems by adapting a small experiment from his most important book.[6] Pull out a piece of paper and a pen, and write the lower case letter "a" ten times, trying as hard as you can to produce identical figures. Now, switch hands, and try the same task with the pen in your other hand. The product of one writer's effort at this task is shown in Fig. 1. The letters could represent any interesting characteristic of a complex system, e.g., results, such as functional status of stroke patients, or temperature in a person with treated osteomyelitis, or satisfaction levels of patients in a health maintenance organization (HMO) pharmacy; or process characteristics, such as waiting times for x-rays, or IV drug dosages dispensed, or depth of incision in a hernia operation; or perhaps characteristics of supplies, such as the tensile strength of suture material, or the level of training of nursing staff, or the memory of a doctor.

The first run of ten letters shows the effects of a system of causes of one general type: common cause in Deming's terminology, random causes in Juran's, or a constant system of causes in Shewhart's. There is variation in the letters, despite the writer's efforts to make them uniform, as there is variation in all measured things. Heart surgery of the type performed on Kim is never done exactly the same way twice, even though the surgical team may strive for such precision. No two doctors have exactly the same set of skills. What makes the letters vary from one another? We can feel certain that there are many causes, e.g., the slope of the paper, a passing puff of wind, irregularities in the flow of ink, and so on. These causes, though in principle understandable, are so numerous and subtle in their influences that their combined effects yield a mathematically orderly, and, in the longer run, predictable distribution of results.

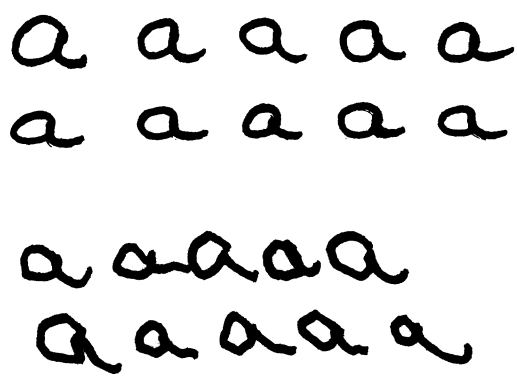

FIG. 1. Two types of variation: Special cause and common cause. The first ten letters were written with the right hand, the second ten with the left hand. Each series contains variation caused by factors *within* the system that produced it (common cause). The difference *between* the two series is attributable to a cause *outside* each system (special cause). Understanding and distinguishing between these two types of variation is basic to modern quality control.

One can never know for certain exactly how large, how round, or how legible the next, as yet unproduced, letter "a" will be, but this series of varying letter forms somehow contains information about its own future. So long as nothing new develops, one can predict, with some confidence, that the next letter "a" will probably be between say,

1.8 and 2.4 millimeters tall and that it will approach roundness within certain limits, and so forth. The series contains information about its own probable future.

Now compare the two series to each other. Variation between them is present, but it is variation of a different type. We happen to know what factor separates them (because we have knowledge of the system of production), but the less obvious fact is that the shapes of the second series contain little information for prediction about the first and vice versa. If one did not know the reason for the differences, the cause would be, in Shewhart's terms, still a "chance cause," that is, unknown; but it is a chance cause of a fundamentally different nature because it does not assist prediction of the future. If letters "a" came generally from series 1 but occasionally and without warning or our knowledge from series 2, then one's ability to predict future forms and sizes would be seriously impaired.

The Value of "Statistical Control"

Those who prepare Kim for surgery rely for quality upon the predictability of the systems of cause that affect her care and her outcome. The surgeon knows that coagulation test reports will be returned within 20 and 25 minutes of their being sent; the anesthesiologist knows that blood gas values will be back in 4 to 8 minutes; the pump technician knows that tubing connections will tolerate pressures in a certain range. Each makes plans in accordance with those predictions, and each bases those predictions on prior experience, which is judged to be informative. Sudden, unpredicted variation is experienced as trouble: Where is the lab result?; or suddenly, the tubing comes loose at a usually safe pressure. These people have perhaps never reflected on Shewhart's formal observation that some variation enables prediction, while other variation confounds it; but they know how each type of variation feels, and they know why the distinction matters.

Shewhart noticed that assignable causes of variation could be discovered and often eliminated from processes if their presence could be detected. He asserted that removing assignable causes of variation, thus rendering processes predictable in their effects, was desirable for several reasons. First, once all assignable causes were discovered and isolated, processes, by definition, behaved predictably in a statistical sense; that is, they could be characterized by such parameters as means and variances. Such processes could therefore be studied for their performance characteristics. If that performance were judged to be unsatisfactory, those controlled processes could be improved through systematic redesign. In other words, Shewhart found a deep relationship between removing assignable causes of variation on the one hand and learning from past experience on the other. For him, statistical control and the scientific method were inseparable.

Second, Shewhart reasoned that to act repeatedly as if attributable causes were present, i.e., ringing false alarms in response to mathematically lawful chance variation, wasted an organization's effort. A physician would be wasting effort if he or she altered a medication dosage based on random fluctuations in temperature. He or she must know when a temperature elevation "means something," that is, has an attributable cause, and then—and only then—act. The same applies to managers faced with fluctuations in the performance data they review.

Third, when processes are in statistical control, that is, behaving predictably, overall costs of production tend to be lower. Less inspection is needed because small or infrequent samples can be used with greater confidence to characterize the lots from which they are drawn. Final products of systems in statistical control are more reliable because their component parts are more reliable.

Shewhart's views of variation, its control, and its connection to quality proved so powerful in improving quality and reducing costs that they now provide the scientific

backbone of most modern industrial approaches to quality management. After Shewhart, understanding and controlling variation became the main mission of the industrial quality professional.

Controlling Variation: A Threat or a Promise?

Yet, the notion of controlling variation strikes fear into health care professionals. The objections come fast: "This matter of quality control may be fine for manufacturing, but I am a physician. I don't make widgets"; or, "Medicine isn't like making cars. The product is not uniform. Every patient is different." Controlling variation, professionals may think, will drive the art, the poetry, and the judgment out of medicine. To the physician, especially, the word "variation" connotes "freedom," and the word "control" connotes "handcuffs."

It is essential that these connotations change if health care is going to benefit fully from modern total quality management. Some of the industrial theory of quality management is negotiable. Compensation policy, for example, is so vaguely developed in industrial quality management that health care organizations can safely experiment broadly with their own compensation policies and then search out those that will most favorably affect quality improvement. However, for anyone who claims to correctly use the lessons of quality management that began with Shewhart, the concern with variation is not negotiable. It cannot be left out of the picture and still leave quality management as a strategy intact.

Why is the understanding and control of variation so central to improving quality? The answer, simply put, is that variation is a thief. It robs from processes, products, and services the qualities that they are intended to have. Variation is in processes what heat is in mechanical systems: evidence of wasted energy. Variation in processes is what entropy is in thermodynamic systems: evidence of the loss of information and of the confounding of prediction.

Distinguishing Intended from Unintended Variation

Physicians who fear an effort to control variation are worried about the loss of options and about insults to judgment. The variation they defend is intended variation, planned variation, introduced for a reason, guided variation. But, this is not the variation that concerned Shewhart; nor is it the variation upon which those who would improve health care quality must declare war. The enemy is not considered, intentional variation, but rather unintended or misinterpreted variation in the work of health care.

Unintended variation is stealing health care blind today. In controlling it, the health care system could potentially recover a bounty in wasted resources that would dwarf the puny rewards of cost-containment to date. It could make rationing unnecessary.

Where does variation enter into medical work? Everywhere.

Understanding Variation in Clinical Data

Think first of the patients themselves and of the tasks of the clinician. Brian was a 16-year-old patient admitted to the hospital with possible osteomyelitis. It was only "possible" because, although the clinical picture and a bone scan in an outlying hospital were consistent with the diagnosis, no organism had been recovered from Brian's bloodstream. Antibiotic therapy had been started on an empirical basis, but Brian had continued to spike fevers for a week after treatment began. He was transferred for further evaluation. The clinical question of greatest importance was this: Did Brian, indeed, have osteomyelitis, with an organism sensitive to the current antibiotic or was a different process operating, perhaps osteomyelitis with a resistant organism or maybe another disease, such as lymphoma?

The diagnostic strategy included careful observation. Over the next 14 days, Brian

was, indeed, observed, and among the observations made were measurements of his temperatures. During that period, his antibiotic regimen was changed three times, he underwent multiple imaging tests, and had both a bone biopsy and a bone marrow biopsy. During those 14 days, Brian had 100 separate temperature measurements recorded in his chart. Those 100 measurements appeared, in fact, on 22 separate pages of nursing notes. Figure 2 shows them in tabular form, although nowhere in Brian's medical record were they shown in just this way.

Show this list to Walter Shewhart, and he would feel quite at home. A measurement system exists, which reports on an important process variable and is placed at the disposal of "operators" (in Shewhart's language) who are to make adjustments based on the measurement. In a manufacturing process, the adjustments would involve dials and levers; here they involve modifications of antibiotics and testing strategies.

When Shewhart studied systems like this at Bell Telephone Laboratories, he discovered that the information was not being used very well. The "operators" of the gauges and machines in fact varied greatly in the ways in which they responded to the information. They varied among themselves, and even a single decision maker varied over time in his or her own apparent rules of action. Operators often overreacted, making adjustments in settings in response to variation that, through the lens of Shewhart's statistical understanding, was simply random. In overadjusting, they produced more variation than they started with. They actually made the system less reliable, instead of more reliable, an effect that Deming was later to call "tampering" but that Shewhart simply called "errors of Type I." Managers, too, tampered. Unable to understand the underlying causes of the variation they saw, managers changed systems in response to variations that were merely random or not caused by the system in the first place, thereby adding complexity but doing no good. Systems got more and more complex, costs rose, and quality suffered.

Does this sound like a modern hospital or not? What are the rules of action that allow a group of six house officers and five consultants to adjust antibiotic dosages based on a stream of 101 temperature measurements? Based upon what statistical theory do they work? Are the changes in management, e.g., hold the antibiotics, start the antibiotics, change the antibiotics, draw a new culture, biopsy the bone, biopsy the marrow, fight the fever with acetaminophen, observe the fever without acetaminophen, systematic interventions on meaningful variations clearly interpreted; or do the clinicians, too, tamper by misinterpreting the signals as noise or the noise as signals? How much of the effort that is poured into the patient, how much of the money, would Shewhart show to be waste, waste that is exactly equivalent to waste the machine tool operator makes when, standing before his or her gauges, he or she adjusts lever after lever in response to meaningless, random, common cause variation?

How much tampering of this exact kind, the kind Shewhart noticed and set about to help others notice, eats into the day-to-day work of clinical management in medical care? No one really knows. The cost could be enormous. Clinicians, flooded today with the results of measurement upon measurement, undoubtedly face serious risks of misunderstanding variation in what is being measured.

Think about the ramifications. Where do clinicians measure and respond clinically based on that measurement? The list is endless. Measure prothrombin times and change anticoagulants. Measure oxygen tensions and change respirator settings. Measure fever and change antibiotics. Measure blood pressure and change antihypertensive. Measure leukocytes and change chemotherapies. Measure pain and change analgesia. Measure electrolytes and change

DATE	TEMPS	Notes/Comments
10/16/90	37.1	OXACILLIN
	39.9	
	38.5	B/C;TYLENOL
	38.1	
	38.8	TYLENOL
10/17/90	37.3	
	37.4	
	37.7	
	37.5	BONE SCAN
10/18/90	38.3	
	37.4	
10/19/90	39.8	OFF ANTIBIOTICS
	37.6	
	36.6	
	37.7	
	38.7	TYLENOL
	38.7	
	39.1	
	37.8	
	37.5	
10/20/90	36.5	
	36.5	BONE BIOPSY
	36.8	
	37.1	
	36.7	
10/21/90	38.2	
	38.2	TYLENOL
	38.4	
	37.0	
	38.0	TYLENOL
	36.9	
	37.7	
	37.9	
	38.7	TYLENOL
	38.8	
	38.4	
	37.1	
	36.7	
	37.0	TYLENOL
10/22/90	36.0	
	37.2	
	38.2	TYLENOL
	38.5	
	38.8	
	38.3	TYLENOL
	38.1	
	37.3	
	37.7	TYLENOL
10/23/09	36.1	
	36.5	
	36.9	
	38.2	TYLENOL
	38.6	
	38.8	
	38.6	TYLENOL
	37.6	
	36.9	
10/24/90	37.8	TYLENOL
	36.6	
	37.8	
	38.4	
	38.4	TYLENOL
	38.7	
	38.9	
	38.5	
	38.4	TYLENOL
	37.8	
	36.1	
10/25/90	36.9	CEFTAZOLIN
	37.0	
	37.0	BONE MARROW
	37.0	
	38.7	
	36.2	
	37.0	
10/26/90	38.5	TYLENOL
	37.1	
	37.3	
	37.5	
	36.5	
10/27/90	36.8	
	36.1	
	37.7	
	38.5	
	38.5	TYLENOL
	37.5	
	37.0	
	36.3	
10/28/90	36.7	
	36.8	
	37.3	
	38.8	
	37.9	
	37.3	
10/29/90	36.1	
	36.6	
	36.6	
	36.9	
	36.9	
	37.0	

FIG. 2. Clinical data for decisionmaking. These 100 temperatures were written on 22 different pages of nursing notes. Interpretations of the signals in these measurements were the basis of multiple clinical decisions. No formal graphical or statistical tools were employed.

IV fluids. Measure and change, measure and change.

Medical practice would be, in this regard, familiar to Shewhart, and yet it has remained mostly unguided by his statistical insights in the day-to-day approaches to managing clinical data. Weed identified this problem over two decades ago, apparently without ever having encountered Shewhart's work; he framed it in terms of the design of medical record systems that could potentially "guide and teach" clinicians.[9] But, in fact, the process of managing data in health care has not changed much at all, even as the volume and complexity of those data have grown by orders of magnitude during this century.

Physicians need not be frightened of trying to master an understanding of variation in these terms. It is only the science that they wish, anyway, to pursue, the science of inference now equipped with more formal tools. It should be fun.

Unintended Variation in Processes of Work

The variation discussed above has to do with individual patients. It emerges in a slightly different form in the context of work

processes. Here, the first trick is to differentiate intended from unintended variation and to notice how costly the latter is.

I had a difficult decision to make about Maria, aged 6. She was thin and had been complaining for months about abdominal pain. Her mother told me that she thought possibly that she had seen some blood in her stool that morning. That called into the picture a number of possible diagnoses, included inflammatory bowel disease, that would require further studies and a specialty consultation. I decided to do a rectal examination on Maria, in part to obtain an immediate stool sample to test for blood. This was not a simple choice because for a young child such an examination is an ordeal that she would not soon forget. Nonetheless, I judged the urgency of the matter sufficient to warrant the discomfort.

With a sample of stool on my gloved finger, I left the examining room to do the test for blood in stool: put a small smear of stool on a paper card, and add peroxide developer fluid. A blue color would indicate blood; its absence would not. Based on the result, I planned either to pursue further diagnostic tests or to temporize.

What I found in the utility room surprised me. For 15 years, when I had gone looking for a stool blood testing kit, I had expected to find a particular type of slide upon which to perform the test: a cardboard card colored green. Green cards had always been there before. But not this time. Instead, suddenly, the cards were red, not green. I turned to the nurse practitioner, who happened to be next to me, and asked what was up. She told me that a new testing system has been brought in: we use red cards now.

That seemed fine, until I noticed that the developer fluid bottle, unlike the red cards, was quite familiar. It was the fluid I had always used with the now-defunct green cards.

"Can I still use the green developer fluid?" I asked.

"Sure," said the nurse. "It works fine."

"Not on your life," said the clinical supervisor who happened to be walking past at that moment. "The new cards take a new fluid. But we're out of it."

"No," said the nurse, "I'm sure the old stuff works."

I headed up to the laboratory, with my glove on, and Maria's tiny stool specimen still perched on the tip of my finger. I found the head laboratory technician.

"There is no green developer fluid down there," she said. "We had all those bottles replaced months ago with the red ones. The green ones don't work on the red slides."

"But this green bottle was in the red box," I told her. "I found it there myself."

"There's always somebody who doesn't get the message," she muttered . "Here's a new red bottle. Please throw out green ones you see down there."

Still with my glove on, I ran downstairs again with the precious red bottle and did the test on a red slide. No blood. I sent Maria home, and 18 months later she is still well.

I should mention, by the way, that I decided to do a little test later that afternoon. I took an old specimen and smeared it on both types of slide, red and green, and developed each with both types of fluid, red and green. It didn't make any difference. Either fluid worked on either slide.

The total cost of the mixup was high, but difficult to calculate. It included twenty minutes of my time, discarded fluid, a lengthy wait for Maria and her mother, and much more. The laboratory supervisor later scolded the pediatrics supervisor, who probably scolded someone else, and so on. Only one thing is certain about the costs of the episode: every nickel of it was waste. None added value, and most of it reflected failed quality. Variation in that work process was the analogue of heat in a mechanical system.

That wasted energy is present every day in health care systems: people intend to do useful work, but the friction in process creates only useless heat. The surgeon turns to the nurse expecting a particular instru-

ment, but there is none on the tray. The operation is supposed to start at 7:00 A.M., but it starts, instead, at 8:00 A.M. The key blood test is drawn from Mrs. Jones, who has terrible veins, but someone used the wrong tube. The x-ray must be retaken. The instructions never arrived, the laboratory report never arrived, the referral slip never arrived, the otoscope bulbs never arrived, the patient never arrived, the doctor is running late.

When people become used to this friction in processes, in Juran's eloquent term, they "disconnect the alarm systems."[10] Flaw is expected; it is no longer a surprise; it no longer even offends us, and we take offense when others less used to it complain. Worst of all, the effort to do away with it ceases because in their hearts people do not believe that it is possible to do away with it. Sometimes, medical people blame the very patients whom they mean to help. The patients seem unreasonable because they expect care processes to work without friction.

All of this waste comes from variation that we do not intend, but that is built, nonetheless, into the processes that we are alleged to control. It is obvious waste. But, it is important also to realize and to admit that some of the waste that health care professionals have learned to accept comes from variation that they do, indeed, intend but that, upon further reflection, they probably should not intend.

Some of the variation that has been built consciously into the processes of medical work is, simply, unwise. Perhaps there are eight orthopedic surgeons who do hip replacements in a particular hospital, and five supply room technicians who prepare the instrument trays. Each surgeon has a special set of instructions for set-up. The supply room technicians talk about "Dr. Mather's tray," "Dr. Sloan's tray," and "Dr. Wilkerson's tray." It has, in fact, become rather a mark of pride in the supply room to get the trays to the right surgeon.

This is variation, and it is intended. But is it wise? The supply room technicians know by now that they must maintain an inventory for each surgeon separately, and they order extra instruments so that each will be satisfied separately. The instruction book for hip replacement trays has 24 pages, three per surgeon, and each surgeon's office has to send an update every year. There are 40 dyadic relationships maintained, 40 pairs of surgeon and supply room technician, and in each dyad there is an investment in learning and a risk of misunderstanding. Perhaps something valuable is being supported with this investment, e.g., pride, autonomy, and customization for the surgeons. But at what cost and with what added risk from the added complexity?

Complexity and reliability are inversely related. Imagine a process with essential components, each of which functions properly 99% of the time. If the process has 10 such components, it will function correctly 9 times out of 10. With 100 components, it will function correctly about 4 times out of 10. With 1000 components, it will function correctly only 4 times out of 100,000.

Both physicians and health care managers tend entirely to underestimate the cost and risk introduced by the complexity that they have built into the processes they maintain. It is a simple statistical principle: the variation in the combined effect of multiple causes grows quickly with the variation in each of those separate causes. Simple systems are more reliable than complex ones.

The notion applies widely. Wennberg has shown the degree of variation in practice that exists among states, cities, hospitals, and physicians in this country.[11-12] Professionals tend to defend that variation for the autonomy it represents. Perhaps, however, they would be better advised to change rather than defend it because of the cost and risk that it creates. The need is not so much for payers and regulators to force the medical system into uniformity—that would be a sad mistake—but rather for the profession and its leaders to recognize that there is em-

bedded in this cacophony of practice so much waste and hazard that physicians simply owe it to themselves and to their patients to reduce the variation wherever they can.

Unintended Variation in Community Systems: The Fragmentation of Purpose

The variation that confounds smooth work within medical organizations even more dramatically affects the processes that work across organizations within communities.

An elderly woman with dementia is evaluated for thyroid disease in an ambulatory clinic, then again by a private doctor, then again upon admission to a nursing home, then again in a hospital when she has about of pneumonia, then, incredibly, again when she is returned to the same nursing home. This is a true story; in fact, eight separate thyroid function workups occurred in 12 months. Don't worry; Medicare paid.

A 12-year-old boy is discovered to have a rare abnormality on his electroencephalogram (EEG) after an episode of head trauma. Does the abnormality indicate structural damage? Six years earlier, during a bout of meningitis, he had had an EEG at another hospital, but after a week of his current hospitalization, no one has yet called the other hospital for the results of the earlier EEG.

We cut funding for alcohol rehabilitation and experiment with liver transplants for cirrhosis. Smoking cessation programs are not a covered benefit, but lung cancer chemotherapy, of course, is. We will not send a social worker to your home to bring you prenatal counseling, but we will resuscitate your severely premature newborn and later treat his bronchopulmonary dysplasia.

Healthy industries change themselves as the needs of the society they serve change. That is how they stay in business. Indeed, in all respects, "the capacity for adaptation" is not a bad definition of the word "health." But to adapt requires the willingness and ability to repeatedly think above old ways and beyond old forms. It involves the willingness to remake things in the service of higher purposes. In this, we are not as courageous today in health care as we need to be. As Senge puts it in his new book, *The Fifth Discipline*: "From a very early age, we are taught to break apart problems, to fragment the world. This apparently makes complex tasks and subjects more manageable, but we pay a hidden, enormous price. We can no longer see the consequence of our actions; we lose our intrinsic sense of connection to the larger whole."[13]

For health care, the "larger whole" is not vague — it is clear. We are here to preserve, restore, and improve health. That is our purpose. It is an accident that we have such forms today as hospitals, encounters, admissions, operations, doctors, nurses, physical therapists, and 50-minute hours. These are fragments, not the whole. They have roots in history but not necessarily in needs.

We have gained a great deal from the fragmentation of the problem of health. The fragments are often excellent in their special roles. We have created traditions and organizational pride and jobs for many. But we have also created our own forms of variation and waste, which occur when the overarching need, the larger whole, i.e., health, runs afoul of the order of the forms. Patients are lost every day now in the transitions among institutions and sites of care; information is lost in the gaps between primary care and specialists and among specialists, themselves. Hospitals maximize revenues as if that, and not health, is what they exist to produce. Without managing care as a system, we can create local excellence and systemic garbage at exactly the same time: locally proud; globally shamed.

The Role of Research

In some important respects, the health services research community has anticipated the pressing need for the health care industry to reduce wasteful forms of variation. Researchers such as Wennberg,[11–12] Brook,[12] Chassin,[15] and others, not managers or pro-

fessional societies, developed our current understanding of the scope and probable consequences of variability in clinical practice patterns. Eddy[16] has dramatized the role of uncertainty in variability among decision-makers and argued vigorously for strengthening the role of both scientific knowledge and decision-analytic rigor in clinical work.

However, in other areas, significant health services research contributions in the science of variation as Shewhart understood it lie not behind us but ahead. The basic principles of statistical process control are not used in practice settings. Clinicians still rely mainly on impression to interpret streams of quantitative information. Impression may, indeed, suffice sometimes, but sound research might well be able to demonstrate that the methods taught by Shewhart and others could have a valuable role in assisting in the interpretation of clinical data and in guiding parsimonious medical intervention.

Little research exists to date on variability and unreliability within processes of care, as opposed to variability in decisions. In industrial quality control, the term "process" refers to the entire system of production in which work occurs. Decisions are only one source of variability in processes; others include, for example, equipment, environment, rules, and measurement methods. It has been true in other industries, and it may prove true in medical care, that the bulk of variation in the performance of processes may be attributable to causes other than the decisions people knowingly make. If so, then health care quality may make its most significant strides not in the control and reorientation of individual persons but rather in the control and redesign of processes of work. To accomplish this, however, we must first gain knowledge of why work processes fail in health care, without simply assuming that it must of course be the people who are letting us down.

Research on the need for systemic reform in health care may be the most challenging and most important of all. Suboptimizing the components of a generally irrational health care system leads to high costs and low reliability. It is time for some bold experiments on the shape of the health care system, itself.

The opportunities for research on quality management are wide. A few questions worth tackling early are these:

1. Can the proper use of statistical process control theory reduce waste, tampering, and iatrogenic complications in clinical settings, especially in intensive care units where continuous monitoring invites excessive intervention?

2. What is the capability of key clinical and support system work processes in hospitals and group practices? (The study of "process capability" is a formal, statistical undertaking in modern quality management.) Which processes tend to be in statistical control, and which tend not to be?

3. What is the epidemiology of process failures in health care systems? What are the most common root causes of failure? (The answers will of course vary widely among processes, institutions, and types of failure);

4. What is the "cost of poor quality" (waste, rework, complexity, variation, inspection, and losses to patients, for example) in economic terms in health care processes? How can we better measure total costs in health care organizations?

5. Where are major loci of unnecessary complexity and of unintended variation in health care processes?

6. Where does suboptimization occur within health care organizations? Is it among health care organizations? Is it between health care and nonhealth care organizations? How can we better manage health care as a system, optimizing total performance? (These questions have technical, economic, and political dimensions, and all warrant attention.)

Overall, the initial agenda of inquiry might be this: Where does undesirable variation occur in health care? How much does it cost? What are its causes? How best can it be reduced? From the answers to these questions will come other questions well worth

the time and attention of the best of our health services researchers.

Conclusion

Research, alone, will of course be insufficient to improve the actual value of health care systems. Effective change will depend on leaders with the courage to help us confront and eliminate the forms of variation we do not, or should not intend. Leaders must help end the confusion between those forms of professional autonomy that must be preserved and those that common sense requires be abandoned. Leaders must begin to break down the organizational and disciplinary forms that health care has inherited and to build new forms that better meet the needs of those we serve. In the end, leaders must explain, controlling undesired or unwise variation is a route not to bondage but rather to trustworthiness, effectiveness, and reliability.

This need for change, we must remind ourselves, has everything to do with Kim tonight, perched on her medical rocket, placed, sleeping, in the hospital's hands. Those who would help Kim owe her mastery over the work they do. They owe her reliability. They owe it to her that every step they take be guided by knowledge and that every act they intend in her service be faithfully carried out. They owe Kim the sweep of vision that will see her in the context of her family, her community, and her life as a whole; and they owe it to her to place their own actions seamlessly in that context. They owe it to her to change if she needs them to change. If they cannot do these things and if they cannot guarantee to her what she needs, then who among them tomorrow morning will be fool enough to launch?

Acknowledgments

The author thanks Judy Kalman for her help in preparing this manuscript.

References

1. Juran JM, Gryna FM, eds. Juran's Quality Control Handbook. 4th ed. New York: McGraw-Hill, 1988.

2. Deming WE. Out of the Crisis. Cambridge, MA: MIT Center for Applied Engineering Studies, 1986.

3. Walton M. The Deming Management Method. New York: Dodd, Mead & Company, 1986.

4. Berwick DM. Continuous improvement as an ideal in health care. New Engl J Med 1989;320:53.

5. Laffel G, Blumenthal D. The case for using industrial quality management science in health care organizations. JAMA 1989;262:2869.

6. Shewhart WA. Economic Control of Quality of Manufactured Product. New York: D. Van Nostrand Company, Inc., 1931.*

7. Shewhart WA. Statistical Method from the Viewpoint of Quality Control. Washington, DC: Department of Agriculture, 1939.*

8. Grant EL, Leavenworth RS. Statistical Quality Control. 6th ed. New York: McGraw-Hill, 1988.

9. Weed LL. Medical records that guide and teach. New Engl J Med 1968;278:593;652.

10. Juran Institute. Juran on Quality Improvement: Workbook. Wilton, CT: The Juran Institute, 1981:1.

11. Wennberg JE, Gittelsohn A. Variations in medical care among small areas. Sci Amer 1982;246:120.

12. Wennberg JE, Freeman JL, Culp WJ. Are hospital services rationed in New Haven or over-utilized in Boston? Lancet 1987;1:1185.

13. Senge PM. The Fifth Discipline: The Art and Practice of The Learning Organization. New York: Doubleday, 1990.

14. Brook RH, Lohr KN. Efficacy, effectiveness, variations, and quality: boundary-crossing research. Med Care 1985;23:710.

15. Chassin MR, Brook RH, Park RE, et al. Variations in the use of medical and surgical services by the Medicare population. New Engl J Med 1986;314:285.

16. Eddy DM. Variations in physician practice: the role of uncertainty. Health Affairs 1984;3:74.

---

*References 6 and 7 are available in reprinted versions from Quality Press, American Society for Quality Control, 310 West Wisconsin Avenue, Milwaukee, WI 53203.