

Annals of Internal Medicine

© 1998 American College of Physicians

Volume 128(8)

15 April 1998

pp 651-656

Developing and Testing Changes in Delivery of Care

[Academia and Clinic]

Berwick, Donald M. MD, MPP

This paper is also available at <http://www.acponline.org>.

Ann Intern Med. 1998;128:651-656.

From the Institute for Healthcare Improvement, Boston, Massachusetts. For the current author address, see end of text.

Abstract

Improving the daily practice of medicine requires making changes in processes of care. In many circumstances, the most powerful way to make such changes is to conduct small, local tests-Plan-Do-Study-Act (PDSA) cycles-in which one learns from taking action. Learning in these cycles has much in common with learning from prudent clinical work, in which therapies are initiated under close observation and adjustments are made as data and experience accumulate. For many system improvements, PDSA cycles are more appropriate and informative than either formal studies with experimental designs (such as randomized trials) or the mere implementation of changes without reflection or evaluative measurement. Physicians can encourage systemic improvement by endorsing and participating in prudent, local tests of change in their own offices and in the health care organizations in which they work. To do this, they must understand the scientific value and integrity of such small-scale tests.

The patient's problem is recurrent cough. Pulmonary function is normal at the moment, but the patient's symptoms and family history suggest asthma. Three physicians confer.

Physician A suggests inhalant therapy. "It's almost surely asthma," he says. "We should begin treatment and move on to the next patient. The best therapy for asthma is an inhaled steroid.

Case closed."

"Wait a moment," says Physician B. "You can't be so sure that the best therapy is a steroid. beta-adrenergics may be better. We need to enroll this patient in a randomized, double-blind trial to increase our certainty about the best medication. Let's get informed consent and draw lots to assign the patient randomly to a treatment group."

"I've got a better plan than either of you," says Physician C. "Our task is to find out what will help this patient, recognizing our uncertainty about both the diagnosis and the best treatment. Let's run a small-scale test. We'll begin with the assumption that the patient has asthma and, on the basis of the scientific literature, we'll start with inhaled steroids. We'll ask the patient to keep a record of her symptoms and to try steroids for a week, and then we'll review the situation to see

what we've learned. If the steroid helps, we can keep it; if not, we can change the dose or test other options."

We can imagine circumstances in which each of the three physicians sounds wise. If the diagnosis is certain, the therapy is well established, and the risk for error is low, Physician A's approach sounds fine. Why belabor the obvious? Appendicitis? Appendectomy. Case closed. But if the science is uncertain and the risks are high, the best course may be to invest in more scientific knowledge, and Physician B sounds right. Take, for example, the choice of chemotherapeutic agents for leukemia. Before one protocol is established as the best choice, responsible care may involve the enrollment of patients in carefully designed randomized trials. This approach always requires that physicians obtain informed consent and keep the individual patient's needs and wishes foremost, but it includes the growth of general clinical science as major goal. For some questions, formal research designs other than randomized, controlled trials (such as quasi-experimental, case-control, and cohort studies) can provide highly satisfactory answers.

But most often, physician and the patient meet on a terrain that is neither of certainty nor of scientific ignorance, and the use of either large-scale, formal research studies or "point-and-shoot" therapy seems inappropriate. The art of medical care involves a continuing, individualized search in which the physician tries to match incomplete scientific knowledge of disease and treatment with incomplete local knowledge about particular patients at particular times in their lives. In caring for patients, we often do not know the way but find the way, step by step. Thus, Physician C sounds wisest to most persons most of the time.

So it is with improvement in general, whether the subject is caring for an individual patient or finding a better system of care. Between once-and-for-all action and formal research design lies a vast terrain in which steady investigation in the form of small-scale tests of change can be useful. Children learning to ride a bicycle call what they do "practicing." So do physicians giving patient care. In the jargon of quality improvement, this form of learning in action is called the Plan-Do-Study-Act (PDSA) cycle. This cycle, also called the Plan-Do-Check-Act cycle, is attributed to Walter Shewhart, founding theorist of the field of statistical process control (related models can be found in the work of John Dewey and others). Specific tests in cycles of action and reflection are at the heart of modern approaches to improvement in organizations, just as they are at the heart of modern approaches to education and learning [1]. The apparent simplicity of the PDSA cycle is deceptive: The cycle is a sophisticated, demanding way to achieve learning and change in complex systems.

Rationale for Plan-Do-Study-Act Cycles

The strong rationale for the use of PDSA cycles in the process of improvement comes largely from systems theory, which was explored in the first three papers in this series [2-4]. Briefly, a system is a set of interdependent elements interacting to achieve a common purpose. Bicycle riding is a system; so is care of a patient with asthma.

From the viewpoint of prediction, systems are unruly. If you change a single element of a system, you may be surprised by the result [5]. A tiny turn of the handlebars of a new bicycle can leave a young rider sprawling on the asphalt, bewildered by the invisible connections between her hands and her balance. A physician who demands that his "special" suture material be placed regularly on an otherwise standard surgical set-up tray may find that the same suture later appears on other trays, where it is not needed. Complexity mounts and errors occur as complex, nonlinear systems yield irrational effects from separately rational causes. An accumulation of reasonable changes produces an unreasonable mess.

In formal science-planned experiments and tests of hypotheses that physicians are trained to value as investigators and to consume as journal readers-researchers try to mitigate the unruliness of systems by gaining experimental control. Randomization is one way to do this. Random selection spreads messy system noise evenly, allowing the one thing of true interest-the independent variable-to shine its signal clearly through. The larger the sample, the more reliably the noise is evenly assigned and the clearer the signal of effect (if any).

When randomization is not possible, formal science has other ways to handle system noise. Epidemiologic methods can help stratify potential confounding factors, either statistically or by design. Cohort studies can use time trends to advantage, and case-control designs can highlight the effects of exposures. But how do these formal methods for inference from data apply in the world of PDSA cycles?

Suppose that we ask a physician why she has just discontinued penicillin treatment in a patient who developed hives while receiving the drug.

"Because she is allergic to the medicine," the physician replies.

"But you don't really know that," we counter. "You need to prove it. Here's how. Give the patient penicillin on 20 different days, randomly interspersed with 20 days without penicillin. Record the presence or absence of hives each day, draw a two-by-two table, and calculate the Fisher exact test for the null hypothesis."

To a good physician, we would sound silly. There is nothing wrong with the Fisher exact test or with examining the null hypothesis, but in this scenario, these tools are out of place. "I don't have to do all of that," the physician would say. "I'm sure enough, and for this patient we have plenty of good alternative therapies as options."

In drawing her conclusions about which medicines to use, this physician is no less a scientist than we are with our Fisher exact test. She is an inductive scientist, learning constantly from

experience reflected upon and informed by her intimate association with the care of this specific patient. She tries, she observes, and she learns through her own PDSA cycles. Randomized, controlled trials would not help much; on most occasions, they would waste her time.

The PDSA cycles-short-cycle, small-scale tests linked to reflection-are powerful tools for learning in complex systems when the aim is to improve those systems. They are most helpful when inaction seems inappropriate but action without reflection seems unwise. In health care, inaction is inappropriate when practice variation is so great that not all practitioners can be right, when practice departs from scientific knowledge or trusted experience, when a system is performing poorly relative to its potential, or when inaction by some creates an opportunity for action by others with less knowledge or bad motives. Research tells us, for example, that breast-conserving surgery for breast cancer is underused in the United States [6] and that most elderly persons with myocardial infarction do not receive beta-blocker therapy in their post-acute management [7]. These findings call for action: Standing still is not an acceptable response, but how can we act with prudence as we abandon the status quo?

Good physicians use learning cycles frequently as they depart from the status quo in sound clinical work. The use of a medication in an individual patient is best approached as a PDSA cycle. We know from formal experiments how a drug will work in general against a given disease, but we always use that drug with the understanding that its action may differ in particular patients. "Try this," says Physician C, "and stay closely in touch."

Small-scale trials are usually the best approach because systems are unpredictable and their dynamics are nonlinear. Small changes can have large consequences, remote in space and time; this makes closed-form solutions and secure predictions difficult or impossible. In health services research, a corollary has been that efficacy (performance of a treatment or device in the laboratory) is almost always greater than effectiveness (performance of the same treatment or device in the field) [8,9]. Researchers in laboratories take pains to isolate the intervention being studied; practitioners in the field cannot avoid the linkage of that intervention to uncountable additional nonrandom influences.

This is not to deny the enormous contribution of formal experimental science to knowledge. When we are uncertain about the benefit of a particular drug or operation relative to that of its alternatives, no investigational design can yield a more confident conclusion than the randomized, controlled trial. In effect, a randomized, controlled trial is a superb form of "final inspection" of a well-crafted, unitary change.

But the same is not true when we are trying to develop a sound change so that it can later be tested formally or when we are trying to adapt a specific treatment or technology for use in a local setting with its own special conditions. Although bicycle riding is well worked out in theory and its physical mechanics can be described in precise detail, no one can "install" bicycle riding in a

child. The proper mechanism for improving a child's bicycle riding always has been and always will be practice, or PDSA.

Formal experimental studies often have another flaw when inaction is inappropriate: They take too much time. A series of small-scale tests that build knowledge sequentially can yield useful information much sooner than many formal experiments and, over time, can be equally conclusive. Especially in a local setting, this series of tests, linked with simple, annotated plots of data over time, can be highly instructive. It can also be far less disruptive of daily patient care than a typical randomized trial, and it can move the process of testing changes in practice into the realm of formal, quasi-experimental design. The statistical foundations for such sequential tests and time series analyses are simple and well established [10].

Indeed, most learning and improvement in real-world settings comes neither from erratic trial and error nor from randomized experiments but from "PDSA science" (pediatrician Paul Miles in Twin Falls, Idaho, calls it "real-time science"). The PDSA cycle is the minimum level of design and thoughtfulness that should characterize efforts to improve health care.

Take, for example, the work of the Northern New England Cardiovascular Disease Study Group in the late 1980s and early 1990s [11]. Their collaborative project used publicly available data on surgical outcomes to study variations in outcomes and processes of care among participating centers. Physicians played a leadership role in trying to make this variation informative. Round-robin visiting among surgical teams, with linkage to a voluntary database on surgical outcomes and complications, allowed clinical groups to identify promising changes in numerous components of cardiovascular surgery, including patient selection, preoperative preparation, bypass pump management, hemostasis, and anesthesia. This rich array of observations (digested in regular meetings of the group) led to numerous local PDSA cycles, the results of which were reported back to the group as a whole. The overall result, achieved without a randomized trial but through many disciplined cycles of reflection and action, was a 24% decline in mortality from coronary artery bypass graft surgery throughout the region. Had the group tried to do a formal, large-scale, randomized trial, they would still be waiting for substantial improvement in outcomes.

Building Synergy and Measuring Progress

Numerous small cycles of change accumulate into large effects through synergy. An intensive care unit can improve by working separately on respiratory care, medication use, admission policies, patient flow, and family support, all at the same time. Its work on respiratory care might include efforts to improve weaning protocols, sedation, ventilator settings, and respiratory hygiene, all at the same time. One PDSA cycle alone is not enough; systemic improvement requires cumulative and linked changes in an ongoing series of tests. This approach stands in stark contrast to large-scale, once-and-for-all implementation of a grand design. Richard Darman,

former director of the U.S. Office of Management and Budget, put his finger on the problem for governments when he noted the tendency in public policy to always go for the home-run change, the once-and-for-all program that will solve a social problem, without the previous discipline of accumulated small-scale tests of the complex design and its elements in action [12].

Tests of improvement are guided by measurements. System improvement, done by using the PDSA cycle, requires both system-level ("What is my batting average?") and process-level measurement ("Is my swing level? Did I predict that pitch correctly?") An intensive care unit might follow risk-adjusted mortality rates, length of stay, and patient and family satisfaction as system-level measures over time and might use specific measures of respiratory function and time intervals to improve weaning processes. A physician group may track the rate of emergency department visits for asthma among its patients (a system-level measurement) while also studying its rate of prescription of inhaled antiinflammatory agents compared with bronchodilators (a process-level measurement) [4].

Can Physicians Feel Comfortable with Plan-Do-Study-Act Learning?

To embrace PDSA learning in the service of improvement, most physicians need to make changes in their own views of the nature of science and study. Four issues may be particularly troublesome for physicians who encounter PDSA cycles in routine work.

First, physicians may need to adjust their ideas about the nature of "rigor" in making changes in their practice. Physicians who find small-scale PDSA cycles insufficiently rigorous should reflect on the pervasive lack of discipline that has, in part, produced the widespread practice variations that have been well documented by Wennberg and Gittelsohn [13] and others. Today, for example, rates of breast-conserving surgery for breast cancer vary more than 30-fold throughout the United States [14], and rates of cesarean section range from 8% in Green Bay, Wisconsin [15], to more than 40% in some hospital service areas. Routine use of the PDSA cycle in the practice of medicine is a far more rigorous approach than most physicians have actually used in the past to justify changes in their own practices, allowing their individual styles to migrate far apart.

In the search for more rigor, we need not turn every proposal for a change in practice into a major research project. The definition of sufficiently rigorous design varies with the learning task. In most local improvement efforts, large sample sizes and randomized trial designs are wasteful compared with reflection on small, informative samples closely observed. The fact that only a small percentage of the population is allergic to penicillin is almost irrelevant when one patient develops hives after receiving penicillin. One episode is enough evidence from which to draw a useful conclusion. In trying to improve the process of care, wisdom often lies not in accumulating

all of the information that could be used to prove a point but in acquiring only that amount of information necessary to support taking the next step.

For example, take one hospital's efforts to introduce a new approach to teaching patients with asthma how to use their inhalers. An improvement team chose to observe five patients in a row who were receiving the new instructions. By the fifth patient, the team had their findings: The instructions were confusing, and a Spanish-language version was needed. The team did not need a large sample or a randomized trial to reach a conclusion useful enough to suggest a next step. Most process improvement is like that, and physicians who greet every proposed change with a demand for lots of data and a randomized trial will only slow the improvement process.

Second, embracing PDSA cycles as a core approach requires a new attitude toward "failure"; that is, failure of a test to achieve its aim. In the world of PDSA thinking, a failure can be far more valuable in building knowledge than a whole series of successes. It is when a reasonable therapy does not work in a particular patient that we may learn the most about the diagnosis. This idea—that failure is valuable for the information it provides—can meet with resistance among physicians and managers for whom success is the only acceptable result. And for physicians, the concept of failure has an additional emotional burden. It is linked to the idea of letting the patient down or forgetting the guiding principle of doctoring: First of all, do no harm.

The solution lies in the very nature of testing; a test is a change with a safety net under it. The PDSA cycle is an effort to minimize the costs and risks of making a change by creating the tightest possible link between action and reflection. In well-run improvement projects, teams are always asking, "What just happened? What did we just learn?" Physicians who really fear failure should not resist PDSA cycles but should insist on them as the only safe way to make forward progress. A well-run test should be designed so that failure does not harm the patient.

Third, PDSA cycles take time and money; improvement takes investment. In a stressed work environment, it is easy for physicians and others to claim that they do not have the energy or resources to support or participate in, much less initiate, tests of change. In the PDSA cycle, the most vulnerable step is "S"—study—because the discipline to reflect on changes, once made, is hard won as long as persons and organizations fail to classify reflection as productive work. Some physician groups have made a large dent in this problem by "harvesting" meeting times; that is, reassigning ritual meetings with little value to the helpful work of starting and reflecting on testing cycles. Such an emphasis on experience and reflection is consonant with modern views of the very nature of effective professional and organizational learning [16,17].

One important way to reduce the costs of PDSA cycles is to establish and maintain measurements of important performance variables over time. If clinical groups keep regular track of such measurements, they can more easily study the effects of deliberate changes, which reveal themselves in deviations from historical trends and patterns. Then, they will not need to set up a new data system every time they want to test a new change.

Fourth, both physicians and patients can find ethical problems in the notion that "testing changes" should be part of routine care. Should all PDSA cycles require informed consent or be subject to review by human subjects committees? I think not, unless (as is rarely the case) treatment decisions are based on randomization rather than clinical judgment. The real choice is usually not between testing and no testing but between informative variation and uninformative variation. Health services research reveals high levels of practice variation among physicians and organizations and even within the same clinical groups over time. The clinicians in one intensive care unit decided, on the basis of available scientific evidence and expert opinion, to develop and test successive changes in their sedation practices to shorten the time during which patients need ventilators. Before they began their PDSA cycles, variation was widespread, but no learning occurred. The PDSA approach allowed them to discover better approaches on their own and, even more important, to test rapidly in their own work the potential improvements found in the scientific literature. If using PDSA cycles in such a case requires human subjects review, why would the variation that preexisted those cycles not raise issues of formal review and consent? In fact, a culture of improvement founded on the systematic small-scale testing of changes and the tracking of outcomes over time is likely, on the whole, to be safer and more closely watched than one in which only spontaneously evolved habits rule without the benefit of ongoing assessment of results.

How Can Physicians Help Lead Small-Scale Testing?

Many health care organizations today are embracing approaches to improvement that incorporate the development and testing of small-scale changes: PDSA cycles. Physicians play an important role in facilitating or impeding such organizational learning. They may be asked to participate in specific trials: for example, by using a treatment protocol for a limited time, helping to test an alteration in the approach to sedation, or assisting in new forms of interaction with patients and families.

An improvement team in one hospital asked physicians in the intensive care unit to allow families to participate in morning clinical rounds to see whether doing this improved communication patterns and the quality and communication of clinical information. In another hospital, nurses asked obstetricians to allow them to ambulate and bathe low-risk patients during labor to help reduce unnecessary cesarean sections. In a third hospital, physicians were asked to standardize their choice of perioperative prophylactic antibiotics to ensure more timely administration and reduce errors. In each case, the change to be tested was grounded in sound scientific literature and the test was confined to a delimited time period, after which the change would be discarded if it had proven ineffective. Resistant physicians could easily opt out of the test and, most important, the effects of the change were monitored over time with respect not only to primary outcome

variables (such as cost, clinical outcomes, and patient satisfaction) but also to "balancing measures" designed to detect negative side effects [4].

Key in all of these examples is the fact that physicians were not asked to agree to a once-and-for-all change or to accept total implementation of a new approach. Instead, they were asked to accept a small-scale trial, confined in space and time and under close observation, so that learning could occur on the basis of action and its observed effects rather than theory alone. For many clinicians, this type of learning requires a transformation of attitude.

In addition to permitting tests to occur, physicians can help in other ways. They can contribute data for the assessment phase of the PDSA process, they can offer ideas for further trials, and they can help interpret the results of current trials by assisting in review and reflection. A negatively minded physician, on the other hand, can do much to inhibit PDSA cycles and thereby slow the pace of learning. Failing to understand the different roles of randomized trials and PDSA cycles and thus always demanding large sample sizes and formal prospective designs where these are inappropriate and unwieldy is one way to impede organizational learning. Another is to continually raise the issue of physician autonomy and control instead of realizing that, by agreeing to the standardization underlying testing cycles, physicians and others truly gain control of the knowledge that they can use to help patients. A third is to be always critical, explaining why a test should not be done or why a proposed change will not be effective, instead of accepting the responsibility to make positive suggestions for change. Such obstructionism causes inaction and creates a vacuum that invites uninformed changes in care by outsiders in response to financial pressures.

Plan-Do-Study-Act Applications in Office Practice

Once they are convinced of the essential role of small-scale, local tests of change in the search for improvement, physicians can find immediate opportunities for application in their own office practices. "Huddles" with staff and colleagues can produce many ideas worth testing immediately. For example, "Tomorrow, let's try taking registration information in the examining room whenever a natural break occurs in the encounter with the physician instead of having everyone register at the front desk. We can assess whether that reduces waiting times or not." The key step in improving office operations is to establish in everyone's mind that small tests of change are expected; valued; and worth the time spent in planning, doing, and reflecting.

Another simple form of PDSA in the office is an on-the-spot question that any patient can be asked at any time: "How did I just do in meeting your needs?" Or, as one obstetrician asked at the end of a routine visit, "In this visit today, was there anything that my staff or I could have done that would have made it a better, more satisfying encounter for you?" The responses to such questions are grist for immediate change, sometimes beginning with the very next patient.

Plan-Do-Study-Act is a simple model for learning that is almost obvious in its appeal. However, long, hard experience has taught us that it is extremely difficult to bring into a busy work environment. Guiding medical cultures-whether in large, formal organizations or in office practices-to accept and value small-scale tests of change as part of the daily routine and as essential steps in the continuous search for improvement is a daunting task. The alternatives, however, can be worse: to accept an inadequate status quo or to take blind, wholesale stabs at change in complex, nonlinear systems where consequences can be dire and hard to predict. Physicians play a crucial role and have a great stake in the choice of the approach that is taken: status quo, "shot-in-the-dark" changes, or the real-time science of the PDSA cycle.

Requests for Reprints: Donald M. Berwick, MD, MPP, Institute for Healthcare Improvement, 135 Francis Street, Boston, MA 02215.

Current Author Addresses: Dr. Berwick (Series Editor): Institute for Healthcare Improvement, 135 Francis Street, Boston, MA 02215.

Dr. Nolan (Series Editor): Associates in Process Improvement, 1110 Bonifant Street, Suite 420, Silver Spring, MD 20910.

REFERENCES

1. Langley GJ, Nolan KM, Nolan TW, Norman CL, Provost LP. *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*. San Francisco: Jossey-Bass; 1996.
2. Berwick DM, Nolan TW. Physicians as leaders in improving health care: a new series in *Annals of Internal Medicine*. *Ann Intern Med*. 1998;128:289-92.
3. Nolan TW. Understanding medical systems. *Ann Intern Med*. 1998;128:293-8.
4. Nelson EC, Splaine ME, Batalden PB, Plume SK. Building measurement and data collection into medical practice. *Ann Intern Med*. 1998;128:460-6.
5. Senge PM. *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday; 1990.
6. Lazovich DA, White E, Thomas DB, Moe RE. Underutilization of breast-conserving surgery and radiation therapy among women with stage I or II breast cancer. *JAMA*. 1991;266:3433-8.
7. Soumerai SB, McLaughlin TJ, Spiegelman D, Hertzmark E, Thibault G, Goldman L. Adverse outcomes of underuse of beta-blockers in elderly survivors of acute myocardial infarction. *JAMA*. 1997;277:115-21.
8. Cochrane AL. *Effectiveness and Efficiency: Random Reflections on Health Services*. London: The Nuffield Provincial Hospitals Trust; 1972:2.
9. Palmer RH. Definitions and data. In: Greene R, ed. *Assuring Quality in Medical Care: The State of the Art* Cambridge, MA: Ballinger Publishing; 1976:14-5.
10. Wheeler DJ, Chambers DS. *Understanding Statistical Process Control*. Knoxville, TN: Statistical Process Controls Pr; 1986.
11. O'Connor GT, Plume SK, Olmstead EM, Morton JR, Maloney CT, Nugent WC, et al. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. The Northern New England Cardiovascular Disease Study Group. *JAMA*. 1996;275:841-6.
12. Darman R. Riverboat gambling with government. *New York Times Magazine*. 1996;1 December 1:6.
13. Wennberg J, Gittelsohn A. Small area variations in health care delivery *Science*. 1973;182:1102-8.
14. *The Dartmouth Atlas of Health Care*. Chicago: American Hospital Publishing; 1996.
15. Sandmire HF, DeMott RK. The Green Bay cesarean section study. III: Falling cesarean birth rates without a formal curtailment program. *Am J Obstet Gynecol*. 1994;170:1790-9.
16. Schon DA. *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. San Francisco: Jossey-Bass; 1988.
17. Kolb DA. *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall; 1984.