The John Eisenberg Lecture: Health Services Research as a Citizen in Improvement

Arguably the greatest achievement of academic health services research of the last half-century has created its greatest unmet challenge. The achievement is to have documented beyond doubt the widespread defects in health care, even in wealthy systems. The challenge is to discover what we need to know that we do not now know in order to create much more effective systems of care.

Health services research has not yet been sufficiently helpful in meeting the challenge of improving care in part because it has over-constrained both its methods and its favorite topics. The cost of insisting on formal, classical, summative, evaluative experimental designs in an uncertain, poorly understood, nonlinear, system is, unfortunately, to maintain the *status quo*. When the *status quo* is harmful, as health care is today, harm is not a theoretical problem. It is real, and it is indecent. Health services research should become more effectively part of the solution. To do that will require that we enrich our portfolio of methods and broaden our agenda of inquiry. The scientific methods that we need to enhance and dignify in academic settings will combine formal classical methods with some pragmatic, immediate, and in many ways more informative forms of learning and investigation.

THE TURNING POINT

Largely owing to the Institute of Medicine, the stage is now set for this important redirection of health services research. The turning point came in 1998 from the IOM Roundtable on Quality with the publication of its findings that "serious and widespread quality problems occur in small and large communities alike, in all parts of the country, with approximately equal frequency in managed care and fee-for-service systems of care" (Chassin et al. 1998). As a result, the Roundtable asserted, very large numbers of Americans fail to benefit as much as they could, or are actually harmed. To most Americans at the time (although not to health services researchers), this apparently immoderate conclusion from an organization as conservative as the IOM was a surprise.

The IOM Roundtable used three terms to classify the defects: overuse, underuse, and misuse, terms that have now become common parlance among students of quality. Research by The RAND Corporation quantified some of the overuse: 20–40 percent of some surgical procedures are performed on patients who cannot, on scientific grounds, be expected to benefit from them (Leape et al. 1991). Similar findings apply to many drugs, tests, and hospital days. "Underuse," the second category, refers to forms of care that could help people but that people fail to get. Underuse is especially prevalent among those in the "safety-net" system, but even average or wealthy Americans often fail to receive care that could help them (McGlynn et al. 2003). The third term, "misuse," was an odd way to denote mistakes—errors in care that cause avoidable harm to patients.

The Institute of Medicine declared the burden of overuse, underuse, and misuse to be very large, and its newly formed Committee on the Quality of Health Care in America harnessed the expertise of dozens of practitioners and academics to get clearer about exactly what the agenda for a new system might be. This effort culminated in the IOM's March 2001 report, *Crossing the Quality Chasm* (Committee on Quality of Health Care in America, Institute of Medicine 2001).

This report presented a picture of care as it could be, but it first took another little understood but very important step. In summarizing the need for improvement, it did not stop at the dimensions of performance that the Roundtable had focused on: overuse, underuse, and misuse. Drawing on information about patients' experience of emotional, psychological, and spiritual care, the *Chasm* report claimed that patients too often feel helpless in care, confused, and left behind. They may get the right technical care, but their emotional and spiritual needs remain unattended. And so the *Chasm* report called not just for safe and effective care but also for patient-centered care. It made a strong case for putting much more control in patients' hands and for honoring the individual patient's needs, values, and desires.

The Committee expanded the aims for improvement even further by declaring timeliness to be an important quality in health care just as it is in other industries. Finally, IOM identified equity as an important quality aim. Much of the effect of race—the most powerful single predictor of lifespan or health status in the US—on health is mediated by factors outside usual health care, but the IOM, in declaring "equity" to be an aim for improvement, put the challenge of addressing these gaps squarely in the lap of a redesigned health care system.

Thus, in the *Chasm* report, the IOM set forth an expanded agenda for American health care quality: six "Aims for Improvement"—safety,

effectiveness, patient-centeredness, timeliness, efficiency, and equity. To the extent that health services research chooses to be a soldier in reducing harm, these are its marching orders.

## THE IMPORTANCE OF DESIGN: NOT "HOW MUCH?" BUT "HOW?"

The data showing the defects are firm. In some of the best health services research of the past few decades regarding technical care, Wennberg and his colleagues have been uncovering tremendous, nonsensical levels of variability in health care use and costs throughout America (Fisher et al. 2003a, b). The *Dartmouth Atlas*, now regularly summarizing this work, reveals, for example, that the number of Medicare beneficiaries per thousand who are admitted to a hospital in a given year for congestive heart failure varies 600 percent, from about six or seven per thousand up to about 40 per thousand, depending nearly solely on where the beneficiaries happen to reside.

Do these enormous "small-area variations" in utilization reflect differences in need, habits of care, supply-driven decisions, or quality? All may play a role, but it is becoming increasingly plausible that the hospital service areas with low admission rates are also places with integrated systems of care, outreach, and the ability to make sense of the journey of a patient with a chronic illness.

Indeed, health services researchers have generally failed to find any predictable relationship between what we spend on care from region to region in the United States and what we get for our money. Working with databases from Medicare, AHRQ, and others (Institute for Healthcare Improvement 2003), Sir Brian Jarman has developed measurements of the probability that a patient admitted in a hospital will die, adjusted for over 100 variables, including both patient-level characteristics and market-area characteristics. Jarman's "Hospital Standardized Mortality Rate" varies over 400 percent among American hospitals (Figure 1). A cloud of residual statistical uncertainty and unmeasured factors of course surrounds every HSMR estimate, but the observed variation is far greater than the uncertainty in measurement.

Using both Medicare reimbursement data and all-payer data from AHRQ's Healthcare Cost and Utilization Project, Jarman has also found a 500 percent spread in standardized reimbursements paid to the same hospitals for the care of those patients. With a 400 percent difference in the adjusted probability of death and a 500 percent difference in payments across hospitals, the

Figure 1:   Hospital Death Rate (Standardized for Age, Sex, Race, Payer, Admission Source, and Type) vs. Charge Per Admission (Standardized for Age and Diagnosis)—AHRQ (1997) Data
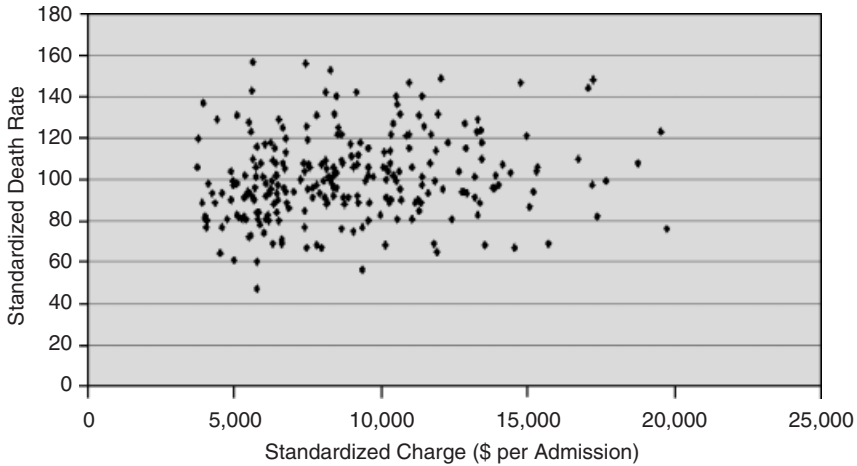
opportunity arises to detect a relationship between cost and outcome. There is none; the correlation coefficient between standardized mortality rates and standardized charges is 0.00. In hospital care, at American levels of reimbursement, more money does not buy better outcomes. Outcomes depend far more on how care is designed than on how much care is given. In short, American health care is irrational, wasteful, and unreliable.

One can quibble (and many do) with the work of Wennberg and Jarman, and with reports in thousands of other journal papers that address similar issues. But in the end, these data draw a disturbing picture of American medicine; namely, (a) it is highly variable, (b) it exhibits very little relationship between the levels of expense and the quality of care, and (c) measurable, predictable, serious defects abound in all six IOM dimensions.

A DUTY FOR HEALTH SERVICES RESEARCH

In my view, this prevailing high level of unreliability and poor quality has, or should have, profound implications for the agenda of academic health services research. Defects in our care create a duty for health services research, both an

opportunity and an obligation, to prevent and reduce suffering. I will explore how to do that, but first I need to comment a bit on methodologies.

The great teachers in the epistemology of modern medicine and clinical evaluative studies, giants like Tom Chalmers, Alvin Feinstein, and Fred Mosteller, brought entirely new levels of importance and excellence to clinical research designs. Through John Eisenberg and others, we learned about how better to connect formal evaluative methods to the work we do as clinicians. As a consequence, we now have embedded in health care an extraordinarily powerful belief system and a set of behaviors around clinical evaluative science. This has taken us a long way from clinical practice guided by anecdote.

Among other consequences, this revolution in applied methods placed the randomized clinical trial at the top rung of design, as the best way to learn. Reinvigorated clinical epidemiology taught us about the hazards of bias and confounding, and ways to control for them. It brought with it sophisticated statistical modeling [methods], and it built our evaluation capacities brilliantly. When hypotheses are clear; when alternatives are discrete and prearranged; when local knowledge is not just irrelevant but toxic to learning (because it produces bias); when we know that confounding influences are present but do not know what they are; when the effects we are seeking are small signals embedded in a lot of noise—when these special circumstances obtain, we need to be as formal as Alvin Feinstein ever asked us to be in his plea for proper evaluative methods (Feinstein 1977; Institute of Medicine 1985). If we become sloppy about research designs and statistical analyses, we can mislead ourselves and others, waste time, and put people at unnecessary risk.

But this commitment to sound evaluative science has also created a problem; namely, that the journey we need to take now in seeking better systems of care will not yield to those methods alone. To crack the problem of health systems improvement, we are going to have to be interested, as colleagues in science, in other methods for learning, as we were previously in evolving the "new classical" methods. The formal methods of summative evaluation simply are not relevant when the hypotheses are many and vague; when alternatives need to evolve over time; when local knowledge is relevant and contains perhaps more transferable wisdom than bias; when the confounders are not defects that spoil our learning, but are themselves interesting and comprise the seeds of further progress; and when the effects sought are large enough that we ought not to have a hard time detecting the signal within the noise.

In the pursuit of pure clinical evidence, controlling for the context and attending to formal sampling are key tactics. However, when the task is to make people safer and better off by acting in the real world, evidence has to be contextualized to be useful. When trying to understand the particularities of local systems, local knowledge and learning processes embedded in the local structure matter a great deal. Without that contextual knowledge, the scientific knowledge remains sterile. The importance of context is at the heart of the familiar distinction between "efficacy" (performance under laboratory conditions) and "effectiveness" (performance in the field).

Correctly done, the combination of scientific knowledge and contextual knowledge should help lead to improvements. To contribute their full measure to reducing suffering, scientific and academic leaders will need to attend carefully to the processes that actually translate science into practice, and they will need to deepen their interest in and mastery of issues in the local context.
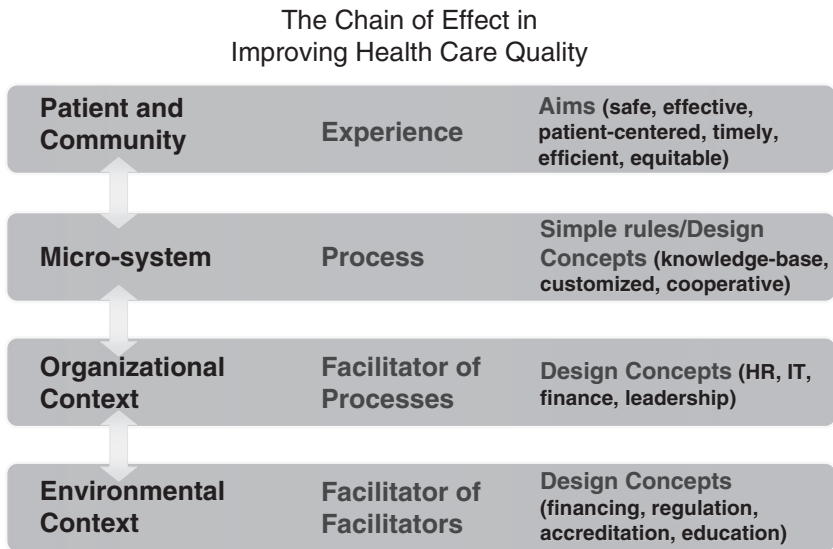
In brief, we need to match the proper form of inquiry to the question. If you want to know for sure if methotrexate is a cure for leukemia, please do a randomized trial. But if you are the Wright brothers trying to figure out how to build an airplane, a randomized trial is the last thing you need. Learning in action is improvement; it is a form of science. We who would improve care are at present much more like the Wright brothers, navigating step-by-step toward better designs, than like Sidney Farber, evaluating a proposed cure for leukemia.

More generally, I urge health services research to take on more fully both the benefits and obligations of a "citizen" enterprise in a total system of work—including nonhealth services research—whose aim is to reduce the burden of illness, injury, and disability among humankind. And if the health services research community accepts this approach, it must attend to both the benefits and the obligations of that citizenship. The question now is this: how do we discharge this duty?

IMPROVING FOUR SYSTEMS

The Institute of Medicine's *Chasm* report argues that the profound changes needed in health care are interlaced among four interdependent systems, nested like Russian dolls, that make up the overall care system (Figure 2). These are (1) the system of patient and community, including the experiences of patients, what we wish those experiences to be, and how our aims relate to this; (2) the microsystems of work, where patients meet clinicians and other helpers; (3) the "macrosystems," especially organizations that support the

Figure 2:   Four Interdependent Systems

The Chain of Effect in
Improving Health Care Quality

Patient and Community	**Experience**	**Aims** (safe, effective, patient-centered, timely, efficient, equitable)
Micro-system	**Process**	**Simple rules/Design Concepts** (knowledge-base, customized, cooperative)
Organizational Context	**Facilitator of Processes**	**Design Concepts** (HR, IT, finance, leadership)
Environmental Context	**Facilitator of Facilitators**	**Design Concepts** (financing, regulation, accreditation, education)

microsystems and create the conditions in which they do their work; and (4) the environmental systems that determine the contingencies that apply to organizations, and, through them to, microsystems.

As citizens in the pursuit of reduction of human suffering, health services researchers have important contributions to make at all four system levels.

The System of Patient and Community

The most crucial contribution to be made at this most basic level is to offer clarity about what is to be achieved. It is easy to think that a system creates aims. However, Deming (1986) suggested the opposite: that aims create a system. That is, until we decide *exactly* what we want to accomplish in the way of reducing human suffering, we cannot specify the relevant system to work or whom to involve.

At this level—supplying plausible, important aims for improvement—health services research has performed brilliantly. Future research should continue to elucidate defects, enhance measurement systems, and explore new frontiers of aim, but the major challenge now lies not in setting aims, but in providing remedies in the form of better information and new designs at the three other system levels: microsystems, macrosystems, and the environment of care.

The Microsystem

The second tier is the microsystem (Batalden et al. 2003). A microsystem in health care is the little unit of work at the interface between care and the people it serves. A microsystem usually has four elements: people working together, like a team in the emergency department; an information system that provides it with the knowledge it uses; a client population, such as the patients who happen to pass through, or an enrolled population (Quinn 1992); and space. The microsystem is the only locus where suffering actually gets relieved; everything else only makes it possible for the microsystem to do its work.

Although health services research has done well in helping to clarify aims for improvement, it has accomplished far less in helping to define approaches for actually improving care at the microsystem level. In fairness, some stellar examples of successful microsystem research do exist, such as the work of Ed Wagner and his colleagues at Group Health Cooperative at Puget Sound and the University of Washington (Wagner 1998). For almost two decades, this group has been tackling the question, "What do we know from science about the systems of care for chronic illness that produce the best results?"

In the mid- and late 1990s, Wagner and his network of researchers around the country digested and analyzed thousands of articles about chronic disease care, and came up with a six-element care model to achieve this productive interaction.

Four of the elements are properties of the care itself: (1) self-management by patients, meaning that patients adjust their own medications, monitor their own physiology, anticipate trouble, and change their lifestyles; (2) design of the delivery system, which is primarily team-based, with a strong role for nursing, including advanced practice nursing, and active outreach; (3) decision support, which puts the right knowledge at the sharp end of care; and (4) clinical information systems, such as integrated, electronic patient records, and reliable disease registries. These four elements of design are housed under the umbrella of the fifth element: an organization that consciously supports, and continually improves, the care system. The sixth element is the connection of care to relevant resources and policies at the community level, such as community-based organizations, employers, and human services sectors other than the medical care system.

This six-element model offers an evidence-based, disciplined, practical understanding of what care should look like. In action, it is brilliantly successful. Dozens of organizations are now using this model to make

profound changes in the well-being, function, risk level, and cost of care for thousands of patients. A recent meta-analysis of empiric studies of the elements of the Chronic Care Model suggests that each element contributes positively to successful patient outcomes, and that they act in synergy, the most important element being self-care (Bodenheimer, Wagner, and Grumbach 2002).

In the resource-poor and disorganized emerging economy conditions of the Russian Federation, Wagner's chronic care model has yielded dramatic results. Using this approach, in Tver Oblast, a Russian "state" with a population of three million, early neonatal mortality decreased from 10.8/1000 to 5.3/1000—a 49.6 percent decrease—in barely a year and a half. Pregnancy-induced hypertension decreased from 43.8 percent of pregnancies to 5.6 percent. Similar results have been obtained in adult hypertension in Tula Oblast (http://www.ihi.org/IHI/Topics/Improvement/ImprovementMethods/Literature/ToRussiawithHealthCareImprovement.htm). The Russian Federation is now extending this model to half of the oblasts in Russia. Researchers from Intermountain Health Care have recently documented a significant decline in readmission rates for patients with congestive heart failure (a relative risk of 0.94, 95 percent CI, 0.90–0.98, for readmission at 1 year) with a coordinated care model similar to Wagner's (Lappe et al. 2004). Overall, health services research on chronic illness care illustrates how helpful systems research can be to the success of microsystems.

Many other important challenges at the microsystem level, by contrast, are not yet studied well enough. If we knew more about problems such as the following, change agents could move faster.

First, we need much more research on appropriate models for understanding the variation in needs of difference subgroups of patients. In its reaction to idiosyncratic and anecdote-based practice, evidence-based health care tends to seek a one-size-fits-all approach to care. We are prone to describing, for example, "the best" way to care for any patient with a heart attack or with diabetes. Some prescriptive models may adjust that care by a few variables, such as age, but we do not yet have other sophisticated ways to stratify our understanding of patients. We should be able to coach organizations to implement, say, "five sizes that fit 80 percent of the population," using robust stratification models that we will change as we learn, so as to achieve what other industries call "mass customization."

Customizing should not be stereotyping, but should rather reflect deep knowledge of the underlying needs of the individual patients served, beyond age, gender, race, diagnosis, and language. This might mean we would know

about patients' particular wishes regarding self-care, their attitudes toward the use of time, their desires for or disinterest in continuity, how informed they wish to be, whether they want the physician to make the choices or want the information to make choices themselves. A sophisticated way of stratifying patients would help us redesign microsystems to be much more responsive and efficient.

Second, at the microsystem level, we need to figure out better ways to incorporate knowledge about variability into the care process itself. At the Children's Hospital in Boston a decade ago, I helped supervise the care of a boy with osteomyelitis of his tibia. He had been in the hospital about 2 weeks and had had almost as many changes in medication and in management patterns as he had had days in the hospital. Working with a medical student, we discovered that he had had 78 temperature measurements in 14 or 15 days, recorded on 72 different pieces of paper, usually embedded in nursing notes and progress notes. The data were all there, but absolutely no learning was taking place from the variation in his physiology.

When we graphed those temperatures, we found that what was being done is what Deming (1986) called tampering; that is, repeatedly introducing instability by intervening in an essentially stable system. Tampering—inappropriate and unnecessary adjustment—is a classic, costly problem in production systems. We do it all the time in health care at tremendous cost (Berwick 1991). Once we understood the random nature of temperature fluctuations in our young patient, we were able to reduce the number of changes in his medications and thus simplify and stabilize his care. . . .

Rigorous, practical use of statistics can help to manage variation better without tampering—to tell the difference between variation that is informative and may require action, and variation that is random and uninformative. Graphical approaches that are far more sophisticated than the methods most health care systems use today are available for understanding variation in the care of individual patients or in patterns of performance in organizations. The related techniques—statistical process control—were breakthroughs in improving work quality for other industries, and applying them would be a breakthrough for health care, as well.

Statistical process control methods have the added power of, in effect, quasi-experimental design. That is, they allow strong statistical inference based on time-series analysis, supporting deepened understanding of cause-and-effect relationships. Indeed, in appropriate circumstances, "SPC" charts provide a more powerful and sensitive basis for causal inference than a randomized trial can. Health services research should thoroughly explore the

potential value of statistical process control methods in health care and clinical evaluative studies.

The third challenge at the microsystem level—perhaps the biggest and most exciting one of all—is to wed engineering sciences more effectively with health care. The beneficial effects of this marriage have been visible for nearly a century at the Mayo Clinic in Rochester, Minnesota. This is not an accident. From the organization's earliest days, the founders of Mayo Clinic managed to bring engineering science to the design of care processes, and Mayo is still doing it. The results are palpable. A patient at Mayo Clinic who is supposed to have a CAT scan at 2:00 P.M., for example, will generally have that CAT scan at 2:00 P.M. Not 1:50, not 2:10, but at 2:00, almost every single time. Most of Mayo Clinic still uses a paper chart (although they are in transition to an electronic patient record), but that chart—famous in the annals of health care record-keeping—is always in exactly the same place, in exactly the same slot in every patient's room. It never fails. It is an extraordinarily reliable system, exquisitely and meticulously designed. And, this highly structured approach to clinical practice occurs at a place well known for its quality of care, contrary to those who argue that well-structured processes are "cookbook medicine."

These are tiny examples of the enormous and exciting challenge of re-introducing formal engineering designs into the systems of care. The cultural and sociological barriers are, however, immense. Health care has made a significant investment in the autonomy of the physician. Unfortunately, over-valuing, even glorifying, that autonomy is toxic to the stability of the systems upon which patients rely. The new era of designing high-quality processes will require a major shift in how health care systems operate, starting with under-standing which forms of independence in physician and nurse behavior are helpful to patients, and which are not.

I am convinced that our system, certainly at its current level of invest-ment and probably at a much lower investment, could achieve wait-free, nearly defect-free, nearly inventory-free care in inpatient settings. But this achievement will require recognizing that, as my colleague Tom Nolan says, this *is* rocket science. The task of scientifically re-engineering health care will not be accomplished simply by a graduate student or a smart doctor figuring out a new schedule on the back of an envelope. It will have to be engaged by our most accomplished health services researchers, cooperating seriously with quantitative scientists from engineering and related disciplines, and engaged for the long haul in real-world settings to test and prove what they discover. Social sciences, like anthropology, sociology, and economics, have much to offer as well, but these have to date been more graciously welcomed

into current health services research efforts than have formal engineering disciplines.

Dr. Eugene Litvak, an engineer, has been studying a big problem in Massachusetts hospitals: ambulance diversion, that is, forcing an ambulance to bypass a nearby hospital because of congestion in its emergency department. In the year 2000, hospitals were on diversion an average of 400 h per month (McManus et al. 2003). Using basic industrial engineering techniques, Dr. Litvak showed, contrary to many observers' intuition, that the correlation is nearly perfect between diversions from the emergency department and the *elective* surgical schedule. What is putting patients on diversion in Massachusetts hospitals most of the time, it seems, is not uncontrollable variation in the needs of patients, but rather potentially controllable variation in the times at which surgeons choose to do elective surgery. Astonishingly, the elective surgery schedule is far *less* predictable than emergency cases.

If Litvak is right (and I believe he is), the bulk of diversions in Massachusetts hospitals could be fixed in one step: by modifying elective surgical schedules and asking surgeons to operate on elective cases when it is technically most appropriate to smooth the flow instead of when they choose. The cultural and managerial issues raised by this intellectual knowledge are of course not dealt with by health services research, but they, too, could be important topics for research. For example, it is likely that implementing scheduling based on Litvak's algorithms would increase hospital net revenues, but working through the macrosystem and environmental issues to implement this change is a daunting challenge worthy of a substantial demonstration study. But before turning to these challenges at other levels, we turn to a final point about microsystems.

The fourth challenge at the microsystem level has to do with the possibilities and limits of self-care. The ultimate vision can be described with a metaphor: "Everyone is a doctor." Most clinical care is simply not that difficult. We could, if we chose to, teach most patients, especially those with chronic illness, to do most of the things most of the time that their doctors actually do. Of course, medical colleagues often rebut this proposal by describing "Mrs. Jones," who is deaf, blind, 84 years old, speaks only German, and has an IQ of 30. "How can she possibly take care of herself?" they ask, to prove their point. Well, the answer is, "She can't." But that is not the question. The question is, "Can the average person take care of him or herself far beyond the levels that today's habits allow them to?" I am convinced that the answer is "yes"—not all the time, but so often that this line of investigation is rich ore for our academic institutions to mine further. Wagner and others have

already shown profoundly improved outcomes when patients with asthma, diabetes, hypertension, or chronic lung disease become far more active agents in their own care. That is just a beginning. We need to set up "skunk works," obstacle-free innovation settings, where we can change rules very boldly, and start with the most capable patients, pushing forward the envelope of possibility for self-care.

The Organizational System

Just as health services research could do more to help redesign the second level of system—microsystems—so it could help more at the third level: the organization. First, and probably foremost, we need a new medical record. Not just an automated record. Not just a slightly modified medical record. It is time to throw the existing medical record away, as Weed (1968) argued so compellingly over 30 years ago, and start again. The medical record of today is fundamentally dysfunctional; yet, unfortunately, it is also so ingrained and so related to the structure of most organizations that "starting again" seems nearly impossible.

The academic community should take the lead in rethinking the medical record from the ground up. We need to ask what we are trying to do by writing things down? How will keeping a record help to relieve suffering? And what properties does that record need to have to be successful? The challenge should be tackled nationally, like a lunar landing. The barriers to change are high; among them are habit, guilt, legal implications, billing, and insurance implications. But somebody has got to start it and this duty could rightly fall to the academicians who study health care systems. A fully redesigned patient record would, for example, have the following features, among others. It would be held, read, and written in by patients as well as clinicians. It would support "rolled up" assessments of performance among patient populations and across settings. It would unify care across locations, disciplines, and specialties. It would support registries and reminder functions. It would embed decision support tools and standardization of care to science, and it would provide such support both to clinicians and to patients. It would be completely legible and always available, and it would favor informative graphs and charts over less informative narratives and paragraphs. In short, as Weed put it decades ago, the record would "guide and teach" all who relied upon it.

I believe that we should create and offer to any clinical office practice and small hospital in the country a thoroughly redesigned, computer-based medical record technology for free. It is a mistake for us to rely on the market

for this. The market will only add variation, which is the last thing we need more of. What we need is a record that generates knowledge at the point of care; in which the quality characteristics of a record system are not only defined but adhered to, foregoing all the bells and whistles that have made medical records two or three times more costly than they need to be. Unfortunately, academic forces have lost traction on this issue.

A second important agenda at the organization level is to redesign the physical spaces where care occurs. Physically, hospitals are not healing places. In fact, good evidence supports the conclusion that medical environments tend to make people sicker—with, for example, excess noise, poor ergonomics, indignities, and isolation. Somewhere in the country, perhaps somewhere that a new facility is to be built, inventors should throw out the old diagrams and the old models and do something that will knock our socks off, something that will create an environment where sick people will not just be cared for, but will heal.

A third important academic challenge in organizational systems is to figure out better how to place technical, clinical knowledge reliably at the point of care. The microsystems cannot solve this for themselves; it is too expensive. A physician who reads one randomized trial per day will be about 10,000 years behind by the end of the first year. The myth of the unaided human (physician) mind as the mainstay of clinical excellence is over. We need to stop relying on the memories of doctors and nurses to assure clinical reliability. Instead, we need to put at their fingertips the best knowledge emerging from the vast enterprise of clinical research, so they can use that knowledge in real time, and so they can spend their time healing, listening, and doing what no knowledge base can ever do.

Putting scientific knowledge reliably at the point of care is not just a matter of compliance and standardization, although it involves elements of both. The academic community has been party to a misperception that the whole job is to create protocols, mandate their use, and ferret out those who refuse to use them. This is a misunderstanding. The first rule in human factors design is this: "Honor thy user." Take people as they are. Create knowledge management portals and systems, including "smart" human bridges to clinical knowledge such as clinical librarian and "informationist" services, that are inescapably easy to use (Davidoff and Florance 2000). We still await a coherent research program that will help us understand what factors in work design—beyond mere incentives—attract and guide physicians and nurses toward scientifically correct practice—what it is that makes the "right" thing the "easy" thing to do.

Another duty of proper health services research is to help place knowledge for improvement in the public domain. In health care redesign, the patenting and owning of knowledge is not a solution. It will not get us where we need to go because the need for knowledge and the need for the ability to judge knowledge are too great. There are potential international solutions that help democratize knowledge: the work of the Cochrane Collaborative and the *British Medical Journal*'s publication of *Clinical Evidence* are hints of what proper, shared knowledge management could look like as a public, global asset.

The fourth challenge for research at the organizational level is to illuminate *muda*—the Japanese term for "waste." American health care critics express great concern about "the business case for quality." I think the business case for quality lies in the concept of *muda*, the Japanese characters which mean, literally, "silly nothing." Silly nothing—*muda*—in a system of production is doing something that cannot possibly help anybody, but doing it anyway.

The needed academic research agenda on *muda* should exploit and mine international variation for learning. The waste levels are phenomenal. In Sweden in 2001, people got care at the rate of about $2,270 per person per year, compared with the US at about $4,887, and outcomes are better in Sweden (World Health Organization). These contrasts in cost and outcomes have little to do with cultural differences; they have a lot to do with design.

This inquiry should begin with a detailed, empiric, multiyear, multinational study, using the very best of process thinking and accounting skills, to understand why, at bottom, health care systems in developed nations other than the US seem to cost so much less than ours. As crucial as this question is, it has still never been taken below broad-scale econometric modeling to a level of detail that would actually help us redesign care for higher value, or at least understand why we cannot do so.

Some years ago, I asked a group of cardiac surgeons to tell me their best idea for improving cardiac surgery. Dr. Steve Gundry, formerly Chief of Cardiac Surgery at Loma Linda University Medical Center, in Loma Linda, California, answered in a single word, "Missions." Then he explained; teams of Loma Linda cardiac surgeons had been going to resource-poor environments in Latin America and Africa to do surgery. There, he said, they discovered what they did *not* need back home. My colleague, Maureen Bisognano, calls this, "Putting on the *muda* glasses." Our health services researchers need *muda* glasses, too. That is a tremendous academic challenge.

A recent study found that the lower the nurse-to-patient ratio, the lower the mortality for patients with acute myocardial infarction (Person et al. 2004).

Interestingly, the study also showed that the LPNs per patient, independent of the RN staff, correlated with *higher* mortality. We need to ask, "Have we now learned what we really need to know by exploring the ideal ratio of nurses to patients?" Or have we avoided a more powerful question from the viewpoint of improvement: "What innovations in work (context, equipment, space, training, care model, etc.) change the association between nurse staffing ratio and AMI outcomes?"

Until health services researchers help us break out of the "more is better" trap, we cannot honor and take full advantage of the work of Wennberg and Fisher showing no correlation at the macrosystem or higher levels between resource intensity and outcomes of care. Factoring in some dependent variables, most related to quality of care, Fisher finds, if anything, a negative correlation between the area-level resources devoted to a condition, such as hip fracture or colorectal cancer, and the risk of death (Fisher et al. 2003b): the more resources expended, the higher the death rate. It appears that Americans may actually pay double in market areas with the highest level of resources. They pay once for those resources themselves, and again for the greater number of complications and other consequences of overuse of care.

One last promising research issue at the organization level involves the question of joy in work. We have no hope of fundamental solution to the American health care dilemma unless it is fun to work in health care. By "fun" I mean satisfying, enriching, fulfilling, and inspiriting. "Joy in work" is in short supply in American health care now, and we know it. The costs of a dispirited workforce are very high. This ought to become a top-priority research issue. We need to understand far more than we do about the psychodynamics of work. How else can we begin to make some scientifically grounded changes that will help our workforce become what they wanted to be when they entered health care: proud, joyful, interdependent actors getting a good job done well?

The Environmental System

This leads us to the last of the four systems, the environment. There are many challenges to be explored at this level and the issues about which methods are appropriate are most obviously complex. To illustrate this system, I will focus on three important facilitators of action at all layers: incentives to provide care, malpractice issues, and workforce considerations.

Health researchers have done quite a brilliant job of giving us ideas about regulatory payment and, to some extent, about the legal issues surrounding the organization of microsystems. But we have not analyzed the

business case properly yet; we do not yet have the cogent frameworks we need to document the toxicities of the current payment system, or how payment systems could better drive recovery of waste and reinvestment of those resources in proper care systems.

We need the help of health services researchers to understand what I would call the "quality characteristics of payment," at a much more sophisticated level than simply adjusting reimbursement to drive incentives. Until health services researchers help us break out of the "more is better" trap, we cannot honor and take full advantage of the work of Wennberg and Fisher showing no correlation at the environmental level between resource intensity and outcomes of care. Factoring in some key intermediate variables, most related to quality of care, Fisher et al. (2003b) finds, if anything, a negative correlation between the area-level resources devoted to a condition, such as hip fracture or colorectal cancer, and the risk of death: the more resources expended, the higher the death rate. It appears that Americans may actually pay double in market areas with the highest level of resources. They pay once for those resources themselves, and again for the greater number of complications and other consequences of overuse of care.

We have a sick malpractice liability system, and it would help if we could get that system straightened out. A proper liability system should be honest and open. People who are hurt should be told about the harm and somehow be compensated for their losses. The health care enterprise, not individual clinicians, should largely bear the liability for this disclosure and compensation. Care systems should commit absolutely to learning from their mistakes, and apology should be, not just possible, but reliable. Today, most clinicians and health care organizations seem incapable of apologizing when someone gets hurt, or too frightened to do so. People who have been hurt in the process of getting health care often say that what they want above all is for someone to say they are sorry. I would like health services research to help us craft some care designs that would help us do so, even while finding ways to improve our ability to remove miscreants and truly incompetent clinicians more reliably and more rapidly from practice unless and until their skills and behaviors are corrected.

Another challenge at the environmental level goes to the heart of work in academic centers: the development of young professionals. Many of the needed changes to improve health care could find ready reflection in the way we teach young people about their role in developing new and better institutional standards, skills, and systems. Young doctors and nurses should emerge from training understanding the values of standardization and the

risks of too great an emphasis on individual autonomy. Health care for the future needs doctors and nurses who understand that cooperation, not heroism, is a primary professional value, and people committed to new norms of transparency, measurement, and continual improvement. Modernizing health care requires a radical redesign in the way we actually teach young people about their jobs in those systems, and none but our academic centers can cause that redesign to occur promptly and properly. Insights and research into solutions to improve value at the environmental level would benefit from international and comparative studies.

ENVOI

The proper redesign of health care as a system, with new and exciting implications for the academic enterprise of health services research, will not be easy to achieve, but neither can it be done bit by bit, without disruption. Trying to solve the shortcomings of the present care system at only one level or another—only at the level of aims and accountabilities, or microsystems, or organizations, or the environment—will not work. The problem is too hard. The Russian dolls are too tightly nested. To raise the bar, as it should be raised, on the relief of suffering, health care will need answers as complex as the system of care itself. This creates for health services research a new level of obligation in its ambition and in its complexity, so that it can take its proper role of citizenship in the very system that it seeks to comprehend.

Donald M. Berwick, MD, MPP, FRCP

REFERENCES

Batalden, P. B., E. C. Nelson, W. H. Edwards, M. M. Godfrey, and J. J. Mohr. 2003. "Microsystems in Health Care: Part 9. Developing Small Clinical Units to Attain Peak Performance." *Joint Commission Journal on Quality and Safety* 29 (11): 575–85.

Berwick, D. 1991. "Controlling Variation in Health Care: A Consultation from Walter Shewhart." *Medical Care* 29: 1212–25.

Bodenheimer, T., E. H. Wagner, and K. Grumbach. 2002. "Improving Primary Care for Patients with Chronic Illness: The Chronic Care Model, Part 2." *Journal of the American Medical Association* 288: 1909–14.

Chassin, M. R., and R. W. Galvin. 1998. "The Urgent Need to Improve Health Care Quality." *Journal of the American Medical Association.* 280: 1000–5.

Committee on Quality of Health Care in America, Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC, USA: National Academy Press.

Davidoff, F., and V. Florance. 2000. "The Informationist: A New Health Profession?" *Annals of Internal Medicine* 132: 996–8.

Deming, W. E. 1986. *Out of the Crisis.* Cambridge, MA: Massachusetts Institute of Technology.

Feinstein, A. R. 1977. *Clinical Biostatistics.* St. Louis, MO: CV Mosby.

Fisher, E. S., D. J. Gottlieb, F. L. Lucas, É. L. Pinder, T. A. Stukel, and D. E. Wennberg. 2003a. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." *Annals of Internal Medicine* 138: 288–98.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and É. L. Pinder. 2003b. "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care." *Annals of Internal Medicine* 138: 273–87.

Institute for Healthcare Improvement. 2003. *Move Your Dot™: Measuring, Evaluating, and Reducing Hospital Mortality Rates.* Boston, MA: Institute for Healthcare Improvement.

Institute of Medicine. 1985. "Assessing Medical Technologies." Report of the Committee for Evaluating Medical Technologies in Clinical Use. Washington, DC: National Academy Press.

Lappe, J. M., J. B. Muhlestein, D. L. Lappe, et al. 2004. "Improvements in 1-Year Cardiovascular Clinical Outcomes Associated with a Hospital-Based Discharge Medication Program." *Annals of Internal Medicine* 141 (6): 446–53.

Leape, L. L., L. H. Hilborne, J. P. Kahan, W. B. Stason, R. E. Park, C. J. Kamberg, and R. H. Brook. 1991. *Coronary Artery Bypass Graft: A Literature Review and Ratings of Appropriateness and Necessity.* Santa Monica, CA: RAND.

McGlynn, E. A., S. M. Asch, J. Adams, et al. 2003. "The Quality of Health Care Delivered to Adults in the United States." *New England Journal of Medicine* 348 (26): 2635–45.

McManus, M. L., M. C. Long, A. Cooper, J. Mandell, D. M. Berwick, M. Pagano, and E. Litvak. 2003. "Variability in Surgical Caseload and Access to Intensive Care Services." *Anesthesiology* 98 (6): 1491–6.

Person, S. D., J. J. Allison, C. I. Kiefe, M. T. Weaver, O. D. Williams, R. M. Centor, and N. W. Weissman. 2004. "Nurse Staffing and Mortality for Medicare Patients with Acute Myocardial Infarction." *Med Care* 42: 4–12.

Quinn, J. B. 1992. *Intelligent Enterprise: A Knowledge and Service Based Paradigm for Industry.* New York: Free Press.

Wagner, E. H. 1998. "Chronic Disease Management: What Will it Take to Improve Care for Chronic Illness?" *Effective Clinical Practice* 1: 2–4.

Weed, L. L. 1968. "Medical Records that Guide and Teach (Part I)." *New England Journal of Medicine* 278: 593–9.

World Health Organization. http://www.who.int/countries/swe/en/.