

INTERVIEW

Wrestling With Variation: An Interview With Jack Wennberg

The creator of modern-day evaluative clinical sciences discusses what motivated him to define and pursue this area of study.

by Fitzhugh Mullan

ABSTRACT: For thirty years Jack Wennberg has studied variations in medical practice, from rates of tonsillectomy in Vermont villages in the 1970s to the cost of dying in the nation's major medical centers today. Along the way he has spawned the field of clinical evaluative science, created the *Dartmouth Atlas of Health Care*, stimulated the creation of a new federal agency (the Agency for Healthcare Research and Quality), and challenged many presumptions about what constitutes good medical care. In this interview with Fitzhugh Mullan, he reflects on health care reform and how to change clinical practice.

Fitzhugh Mullan: You can fairly be credited with being both the Christopher Columbus and the Johnny Appleseed of clinical variation—you discovered it, and you have worked hard to bring it to the attention of the medical and health policy communities. But when you started your work, the common presumption was that doctors practiced in “usual and customary” ways, which were quite well standardized. The term “variation” as you came to use it was really not part of the medical vocabulary. The first time that you wrote about variations in health services, I believe, was the paper you published with Alan Gittelsohn in *Science* in 1973. How did you get started studying variations?

John Wennberg: The early work was done when I was at the University of Vermont in the early 1970s as the director of the state Regional Medical Program (RMP). I had just finished

my medical residency at Johns Hopkins, where I'd trained as an internist with a specialty in renal disease. But I had also taken an MPH [master of public health degree] and started on a doctorate in sociology. The RMP had a large budget with quite a vague set of goals having to do with controlling heart disease, cancer, and stroke. So since I'd been trained in epidemiology and interested in social systems, it was a fairly natural thing for me to want to develop a system for measuring the performance of the system, particularly since our goal was to regionalize care, in an effort to better combat heart disease, cancer, and stroke. We set up a data system and developed a strategy for measuring resource inputs to market areas in an effort to correlate resource inputs with utilization. We wanted to measure outcomes, but mortality was about the only thing immediately available to us.

.....
Jack Wennberg (john.wennberg@dartmouth.edu) directs the Center for the Evaluative Clinical Sciences at Dartmouth Medical School in Hanover, New Hampshire. Fitzhugh Mullan (fmullan@projecthope.org) is a pediatrician, writer, and former director of the Bureau of Health Professions in the U.S. Department of Health and Human Services. He is a contributing editor of Health Affairs and author of Big Doctoring in America: Profiles in Primary Care (University of California Press and Milbank Memorial Fund, 2002).

The whole process was made feasible by Kerr White's work. Kerr had been at the University of Vermont prior to coming to Hopkins, where I was fortunate to study with him. Kerr had persuaded most Vermont hospitals to join a hospital discharge abstract system called the Physician's Activity Study. For every hospitalization, it generated information on the patient's diagnoses, surgical procedures, age, sex, and place of residence. For hospitals that didn't belong to the data service, we sent RMP staff into their record rooms to make our own abstracts. We thus obtained information on virtually all hospitalizations of Vermont residents. We also sent our staff into all nursing homes and home health agencies to obtain similar data. I was fortunate to be able to get the Medicare Part B database by simply going to the Blue Cross offices in Concord. We then began to examine Vermont's health care system from a population-based, epidemiologic perspective. Some people say that we invented the concept of medical care epidemiology in this process.

Mullan: What led you to your observations on variations?

Wennberg: We had a very extensive database, which allowed us to look not only at the acute sector, but at the ambulatory care sector, home health care, and nursing homes. We divided Vermont into local hospital service areas based on how patients in each Vermont town used the system. For each "market" we could view the variation topic as a systems problem; were there trade-offs between use of hospitals, nursing homes, and home health agencies? The short answer was no. And we were able to develop measures to quantify the physician workforce allocated to each market—labor input is what we called it. We knew the number of internists, pediatricians, surgeons, and GPs [general practitioners] in various areas per 1,000 residents and could correlate those numbers with the number of hospitalizations and procedures; we could do the same thing for the numbers of hospital and nursing home beds. What we found was wide variation in resource input, utilization of services, and expenditures among neigh-

boring communities and a strong association between resource inputs and use rates. Inevitably, we were led to the question: Were more doctors and more procedures actually producing better outcomes in terms of mortality?

Mullan: Did you have a sense that substantial variation was out there, or did you discover it as you went along?

Wennberg: I think it was probably the latter, in the sense that there was virtually no literature to suggest the importance of supply in determining use rates. We really didn't know what to expect. I went to Vermont believing in the general paradigm that science was advancing and that it was being translated rationally into effective care. At that time, economists and sociologists as well as patients and doctors believed in the concept that the physician was competent to act as the purchasing agent for the patient—that delegating decision making to the doctor led to wise choices on behalf of the patient. Public policy was also framed around the belief that the physician also acted as an agent for society so that when resources were used to capacity, society could confidently respond by increasing capacity to meet medical need as defined by the doctors. We could thus rely on the "agency" of the doctor for the well-being of the patient and the system. The central tendency of the market was rational.

I had read enough sociology and was aware of the overt and covert functions within systems that I came to the RMP work armed with some skepticism about human behavior. Having read that literature, I was prepared for interpreting what we found. But I don't think I went into it with the expectation that we would find a such a marked variation in medical practice. Variation, as it turned out, was everywhere. For instance, we lived between Stowe and Waterbury. My kids went to the Waterbury school system ten miles down the road. But if we had lived about a hundred yards north, they would have gone to the Stowe school system. In Stowe 70 percent of the kids had their tonsils out by the time they were fifteen years old, as opposed to only 20 percent

in Waterbury.

Mullan: How did your documentation of variation link to your subsequent observations about supplier-induced demand?

Wennberg: Our data suggested supplier-induced demand. The rates of surgical specialty activity were strongly associated with the presence of the respective types of surgeons. More internists were associated with more diagnostic tests and physician visits. And then there was that epidemic of tonsillectomies near my home. It was pretty easy to conclude, from the ground level, that supplier-induced demand was strongly operative.

Mullan: Why did you choose to publish this work in *Science*?

Wennberg: I didn't choose it. We tried the conventional medical journals and received form-letter rejections. This still happens. Generally, we don't bring good news. *Science* was the journal of last resort, but we were delighted to get the paper accepted.

Mullan: What effect did the paper have?

Wennberg: The response was muted initially but gained ground over time. The paper was very important for me because it described a set of problems that have occupied me ever since: What are the causes of unwarranted variation, of variation that cannot be explained on the basis of illness, patients' preferences, or dictates of scientific medicine? What are the consequences? When is more better? When is there too much care and the attendant likelihood of iatrogenic illness? Under what normative standards should variations in health care delivery be evaluated?

Mullan: In 1982 you published "Variations in Medical Care among Small Areas" in *Scientific American*, in which you recapitulated many of these themes. In that piece you also raised the notion of patient preference—the informed consumer. These ideas were a departure from the conventional medical thinking of the time. What led you to them?

Wennberg: I left Vermont in 1973 and went to Harvard. John Bunker, Benjamin Barnes, and Fred Mosteller had organized a yearlong seminar to examine surgical practices. The

group included economists, decision theorists, epidemiologists, biostatisticians, and clinicians all sitting around, haggling, trying to figure out what was going on in surgical practice. The variation issue struck home for these people. Duncan Neuhauser wrote a paper on hernia operations in which he presented evidence that patients' preferences were quite variable. That was the first time I'd seen patient preference as a strong issue.

Mullan: You moved to Dartmouth in 1980, and from that time on your work was closely associated with Maine and with prostate disease. How did that come about?

Wennberg: In 1973 Dan Hanley, who was editor of the *Maine Medical Journal*, published three articles by Alan Gittelsohn and me that showed that Maine suffered as much as Vermont from wide variations in care. But Dan was interested in more than simply publishing the data. With financial support from the Commonwealth Fund, he established a program in Maine to organize physicians to respond to practice variations. The program was organized around what we called the three-step process: (1) Look for an obvious explanation involving "bad behavior" by physicians. (2) If this doesn't resolve the variation, assess the literature and see whether or not it is possible to resolve the uncertainties and conflicts among the different clinical camps with existing information. (3) If this doesn't work, undertake outcomes research.

The group that really took off was the urologists. We had recorded striking variations in surgery for benign prostatic hypertrophy (BPH), a noncancerous enlargement of the prostate. In some parts of Maine, 60 percent of men had their prostates removed by age eighty; in other parts less than 20 percent did. As we went through the three-step process, it quickly became evident that basic facts regarding the outcomes of care for BPH were missing. But even more surprising, the physicians themselves weren't all on the same page when it came to the reasons for doing surgery in the first place. After a prolonged and sometimes heated debate, it became apparent that there were two schools of thinking about why

surgery was indicated. Most worked under the hypothesis that BPH surgery was required to make people live longer—that early surgery prevented development of bladder obstruction, kidney failure, and premature death in later years. But a minority thought that the natural history of untreated BPH was for most men quite benign—the risks associated with early surgery were not paid back by a significant gain in life expectancy. For them, the reason for doing surgery was to improve the quality of life by reducing urinary tract symptoms.

At this time, Al Mulley, Mike Barry, Jack Fowler, and I had been meeting to develop a strategy for conducting outcomes research. The willingness of Hanley’s urologists to join forces with us to undertake the third step in the process, and the generosity of the Hartford Foundation in funding our work for more than a decade, provided the opportunity to put our interdisciplinary strategy into action and to see it through to completion. In brief, the preventive theory was shown to be wrong. The principal reason for doing surgery, it turned out, was to improve quality of life as it was affected by urinary tract symptoms. But it also became clear that BPH surgery could have a negative impact on another important aspect of quality of life: sexual functioning. In other words, the decision to undergo prostate surgery involved a significant trade-off between urinary tract health and sexual functioning.

Coming to this understanding was an important milestone for our research team. We now had concrete evidence that for at least one important example of treatment variation, the key to rational choice for patients and for learning how much resources society would have allocated to meet surgical “need” was to overcome the old model of agency or delegated decision making. Medical ethics and good economics require that patients be actively involved in making decisions. In this context, the concept of shared decision making first

emerged as the principle remedy for unwarranted variations for what we would later come to call “preference-sensitive” treatments.

This understanding led naturally to the next phase of our work: the development of decision aids to help patients sort out the complexity of treatment choices. We took advantage of interactive video technology that was becoming available at that time to develop a program to inform patients. We then undertook to study its impact on decision making.

We learned some very interesting things. First, once patients are informed about what is at stake, most are willing—indeed, anxious—to participate in choice of their own treatment. Second, when patients participate actively, the treatment chosen more closely corresponds to their own values than it does when doctors choose for pa-

tients. We found out that patients who were very concerned about the negative impact of BPH surgery on sexual functioning tended to choose watchful waiting rather than BPH surgery, while those who were very concerned about urinary tract symptoms chose the opposite. Third, we began to get some benchmarks that let us know something about the “right rate” for BPH surgery—the rate that happens when patients rather than suppliers determine the rate. This was possible because one of our experimental sites was a staff-model HMO [health maintenance organization] where we could observe the rates of surgery before and after the implementation of shared decision making. Although the prestudy rates were already low compared to most places in the United States, once shared decision making was introduced, the rates dropped 40 percent, to a rate that was at the very bottom of the national distribution of BPH surgery at that time. The implication seemed pretty clear to us: The amount of BPH surgery provided in most parts of the United States probably exceeds the amount of surgery that informed patients want.

“When patients participate actively, the treatment chosen more closely corresponds to their own values than it does when doctors choose for patients.”

The Politics Of Research

Mullan: You have worked hard to bring quantitative methods to topics that weren't recognized at all or thought not to be the stuff of hard science—practice variation, small-area analysis, patient preference. Yet your thinking seems frequently to challenge the established order, to bring controversy to practices that have been long established and well accepted. Do you see your work as having a political mission?

Wennerg: Since it has so much to say about how health care markets seem to be working, it was relatively easy to attract attention. It really began in 1984. *Health Affairs* published a theme issue on practice variations, and we held a press conference at the Capitol to publicize it. Someone from the AMA spoke, I spoke, and some members of Congress spoke. I think that planted a seed. Some of the politicians were intrigued by variability. Over the years several hearings were held on the Hill. Bill Gradison [R-OH] in the House and David Durenberger [D-MN] in the Senate became interested and supportive. Dan Hanley recruited Sen. George Mitchell [D-ME]. He and I met with Mitchell more than once and persuaded him that he needed to get involved.

Mullan: What did you have in mind?

Wennerg: The initial idea was an amendment to National Center for Health Services Research (NCHSR) legislation to fund variations research. That was quickly supplanted by the more ambitious legislation, which was enacted, that established a new Public Health Service agency, the Agency for Health Care Policy and Research (AHCPR) in 1990. And at that point, I was active politically, working with Congress and also with the research community. I saw this as a winner for everybody. Our goal was to introduce clinical research into the health services research agenda, which had been dominated pretty much by policy wonks and economists. The legislation called for a new research vehicle, Patient Outcome Research Teams (PORTs), which would carry out medical effectiveness research on specified clinical problems asso-

ciated with costly practice variation. The concept was built on the model of research we had developed in Maine, which established an interdisciplinary group of well-qualified researchers to “patrol” a clinical problem such as BPH. The goal of the patrolling was to uncover and explicate theory and to apply various analytic tools and methodologies to test that theory. It was also to keep up with innovation by bring new, promising technologies into early clinical trials as soon as possible. A number of PORTs were launched focusing on coronary artery disease, arthritis of the hip and knee, BPH, and low-back pain—conditions for which one of the treatment options involved discretionary surgery. While a good deal of progress was made in clarifying theory and explicating the role of patients' preferences, the clinical trial networks to develop prospective evaluation on new treatment theories never developed.

Mullan: The birth of AHCPR at the beginning of the 1990s was followed fairly quickly by the health care reform period, during which you were also quite active. As I recall, there was the belief that the evaluative sciences would be at the core of a reformed health system, that outcomes research would be the coin of the new health realm. What happened?

Wennerg: Health care reform turned out to be a disaster. I did get involved because I knew Hillary Clinton through Chick Koop [C. Everett Koop, the former surgeon general]. Chick and I were asked to read the whole reform plan over at the very end, and my task was to be sure that outcomes research and patients' preferences were adequately present. We also were able to define a special role for practicing physicians in governing medical practice by including a professional foundation provision based on Dan Hanley's work. But it all went down. I think it was doomed all along. Health care reform was like the donkey without a tail. Everybody wanted to put their own on. It just didn't work.

Mullan: AHCPR has had a mercurial life in the decade since then, including the acquisi-

tion of a new name: the Agency for Healthcare Research and Quality (AHRQ). What kind of marks do you give it?

Wennberg: In the mid-1990s the agency suffered the coming together of several bad things all at once—the failure of the Clinton legislation, the “Gingrich revolution” in Congress, and a handful of dissident orthopedists and neurosurgeons. This latter group didn’t like the fact that the PORT under Rick Deyo and Jim Weinstein found that there were some serious problems with back surgery that needed national attention. They went directly to their congressmen, who considered the surgeons’ complaints one more reason to zero out a government program. By that time Mitchell, Durenberger, and Gradison had all left Congress. The agency escaped annihilation, but it had to abandon funding for the PORTs and other clinical analyses that sought to improve the scientific basis of medicine by testing conventional medical theories. I think that John Eisenberg saved AHRQ because he was respected and because he picked a safe topic for his agenda: medical errors. I mean, who wants medical errors? Everybody’s against medical errors. Today AHRQ is irrelevant to the problems that I was interested in having it address. It is out of the business of determining the scientific bases of clinical practice. That’s why I believe that we have to establish the concept of clinical outcomes research as a central theme at the NIH [National Institutes of Health].

The Role Of The NIH

Mullan: I interviewed Elias Zerhouni for *Health Affairs* recently (Web Exclusive, 8 January 2004), and I asked him about the role of the health services research at the NIH. He responded with a comment to the effect that you can do research that looks for new science or you can do research that looks at “the difference between Coke and Pepsi.” He said he didn’t feel that both types of work could be done well at the NIH, and he favored the pursuit of new science. As long as your work is viewed as soft drink sampling, it seems

hard to envision it at the NIH.

Wennberg: That’s a mindset that needs to be challenged, especially among people who believe in science. Medical theories need to be tested, and we’re awash in untested theories. Even new technologies that have passed the muster of a clinical trial (and many have not) move into practice in many unevaluated ways. What is done with a new technology once it’s in the market depends on the inventiveness of physicians, and they’re terribly inventive. Untested theories and practices are huge, expensive, and dangerous problems.

Mullan: Why is it so important to move the evaluative sciences onto the NIH campus, and why has the biomedical research community been so reluctant to embrace them?

Wennberg: The evaluative sciences are an important part of biomedical science, a view that is not permeating academic medicine sufficiently at this point. The research community doesn’t pay attention to evaluative science, in part because there’s no funding for it. There’s no basic training to speak of. There are no careers. In bench science or clinical investigation, researchers can get funding to work on a problem for decades. Through the PORT concept, we tried to make the same model work for the evaluative sciences. This is a problem that goes way beyond the research community. The employer community, the tax-paying community, and the Centers for Medicare and Medicaid Services (CMS) need to understand that their cost problems relate, to a large extent, to unevaluated technologies. They need to give political support to a major upgrade for funding for evaluative clinical science. I would advocate moving AHRQ into the NIH, broadening its mandate, and increasing its funding to a billion dollars a year by getting contributions from the CMS and from the insurance industry, so that the agency goes into the NIH without competing for existing NIH research.

Mullan: You don’t think there’s inherent hostility toward health services research at the NIH?

Wennberg: I don’t think so. No scientist would be hostile to this kind of work.

Mullan: But they don't consider it a priority, and some clinicians, as we've seen, aren't eager to have evaluation scientists examining their practices.

Wennberg: I'd beg to differ on that. I think that most clinicians would not take that position. There will always be somebody whose theory is gored by evidence. I mean, that's just the way life is. In the give-and-take of medicine's evolution, there are going to be technologies and strategies that work better than others, and some will have to be weeded out. You just can't live with the old and the new and have a sustainable economy.

The Dartmouth Atlas

Mullan: Tell me about the *Dartmouth Atlas of Health Care*. Where did the idea come from?

Wennberg: It came from all of our work in small-area analysis, which is quite geographical. We wanted to be able to use our data in a variety of different contexts—regulatory contexts as well as clinical management. It was the Clinton health plan that really motivated us, because we anticipated that it would be built around “Health Care Alliances” that were geographically based health insurance areas. So we approached the Robert Wood Johnson Foundation, and they gave us a large grant to take the Medicare data and to organize a national small-area analysis based on methods similar to those we had used in Vermont. When the Clinton legislation crashed, we had a lot of data but no customer. So we published an atlas. It has turned out to serve a lot of useful functions. It keeps reminding people about fundamental imbalances in health care, of the pervasiveness of supplier influence on utilization. The media love it and use it all the time.

The most important evolution is that over the last four years, we've moved beyond variation by geographic area to analyze variation among health care organizations. We can look

at cohorts of patients who use one hospital or another and compare the resource allocation between them. We now know the answers to questions that previously could be asked only of staff-model HMOs. How many doctors per thousand are they using? How many hospital beds? What's their surgery rate? How do they manage chronic illness? How many physician visits do they provide? How many days in the hospital? In intensive care? How much does it cost per capita? So all the variables that have

been available at the area level now have an analogue available at the hospital-specific level.

Hospital-specific information is important because it should stimulate action. We hope that a recent article we published in the *British Medical Journal* will help motivate academic medical centers to get involved in rationalizing their own practice patterns. The variation is truly extraordinary.

For example, during the last six months of life, patients who used the NYU [New York University] teaching hospital made, on average, seventy-six visits to physicians, 57 percent saw ten or more physicians, and the average patient spent almost a month in the hospital; patients using UCLA [University of California, Los Angeles] hospital spent, on average, 9.2 days in intensive care, patients made forty-four visits on average to physicians, and 51 percent saw ten or more physicians. Contrast this to the experience of patients loyal to UCSF [University of California, San Francisco]: The average patient spent less than twelve days in the hospital and only 2.6 days in intensive care; patients made twenty-seven visits to physicians, on average; and only 30 percent saw ten or more physicians.

View Of Personal Success

Mullan: Your writings and teachings over the years have achieved great credibility. Although health care has changed considerably during your career, it has not always changed

“Lurking behind variations in patterns of care are often huge hospital investments in expensive technologies that are directly tied to their economic stability.”

in the direction that your work might have suggested. Variation in practice is still rampant. The penetration of evaluative science into the research world has been limited. U.S. health care remains expensive and not very efficient. Are you frustrated? What is your take on your own level of success?

Wennberg: Well, I guess it just depends on how long one is willing to wait. I think that these issues will eventually reemerge as mainstream. We keep trying to prod it a little bit. I am hopeful that our recent progress in profiling academic medical centers will ignite the debate once again. The significance of these differences must surely be more than the difference between Coke and Pepsi. They are associated with more than a 2.5-fold difference in per capita costs; moreover, as Elliott Fisher and his colleagues have shown, the overuse of these services seems to be associated with worse outcomes.

Mullan: When you say “reemerge,” that implies that there was a period when they were mainstream?

Wennberg: Well, I think we almost had it at the time of the Clinton health care reform. I think it will happen again. The system is stubborn because it is, after all, 15 percent of the economy. There are a lot of oars in that water, and information per se is not going to change the fundamental economic incentives.

We’ve been trying to pursue models of reform that would allow the reimbursement system to align itself with the quality agenda. First, we think that in most parts of the country we probably have an excess capacity of both physicians and hospital beds, in terms of what’s beneficial to the population. Second, we don’t know what the supply of specialists, particularly surgeons, really should be, if patients were the determinants of demand. We know what the supply is now, and we know that it’s fully utilized everywhere, no matter how much there is. We also know that the reason for this is physicians’ influence over decision making. But once you open the market to information, as we’ve shown in several clinical trials of decision aides, demand drops among informed patients. We have identified three

types of unwanted variation in the system—the underuse of effective care, the misuse of preference-sensitive care, and the overuse of supply-sensitive care. Lurking behind variations in patterns of care are often huge hospital investments in expensive technologies that are directly tied to their economic stability. We have proposed the establishment of a Comprehensive Centers of Medical Excellence program in which medical centers would partner with Medicare, AHRQ, and the NIH to develop methods to deal with unwanted variation in the system. We were very pleased to see our proposal become law as part of last year’s Medicare reform bill.

The Inevitability Of Reform

Mullan: A number of times in the mid-1990s I heard (as you perhaps did, too) the late Eli Ginzberg fret about the future. “We have made it to a one-trillion-dollar health care system,” he would say, “but I really don’t think the economy can sustain a two-trillion-dollar system. Something will have to give.” Eli is no longer with us, but we are moving rapidly toward his two-trillion-dollar Armageddon. What do you think? Is there some dollar figure or percentage of gross domestic product that will bring disaster or reform?

Wennberg: I don’t know about the number, but I think the trend is pretty clear. Employers are basically giving up. They’re trying to shift health costs to their employees, or they’re not providing anything at all. Once the employers give up on employer-based health insurance, the demand will grow to add health insurance to the national tax agenda, which all other civilized countries seem to do. That’s when I think reform will happen. It’s happening already. Many health plans now offer “tiered products” in which patients have poorer and poorer coverage—a kind of skin game. But I think there will be only so much skin that people will put into this game before they try to change the rules.

Mullan: Those new rules would provide for less variation and better quality of care at a cost that is sustainable.

Wennberg: That would be my hope.