

Interview with Dr. Donald Berwick

This month, AHLA President Elizabeth Carder-Thompson spoke with Donald M. Berwick, MD, President and Chief Executive Officer of the Institute for Healthcare Improvement, a nonprofit organization seeking improvement of healthcare throughout the world. Their wide-ranging discussion included multiple aspects of comparative effectiveness—a key concept in the current health reform debate—as well as electronic medical records, future formularies, and regional variations in medical practice.

In addition to leading the staff of the Institute, Dr. Berwick is a Clinical Professor of Pediatrics and Health Care Policy in the Department of Pediatrics at the Harvard Medical School and Professor of Health Policy and Management at the Harvard School of Public Health. He graduated with a BA from Harvard College, and received an MPP from the John F. Kennedy School of Government and an MD from the Harvard Medical School. He completed his medical residency in pediatrics at Children's Hospital Boston.

Dr. Berwick has published articles in professional journals on subjects relating to healthcare policy, decision analysis, technology assessment, and healthcare quality management. His research and commentaries have appeared in *The Journal of the American Medical Association*, *The New England Journal of Medicine*, *The British Medical Journal*, and others. Dr. Berwick is a past chair of the Health Services Research Review Study Section of the Agency for Health Care Policy and Research, and a former Chair of the National Advisory Council of the Agency for Healthcare Research and Quality.

Elizabeth Carder-Thompson (ECT): The first grouping of questions is on comparative effectiveness. How do we determine the threshold of evidence that demonstrates that a medical treatment is scientifically based and sufficiently effective?

Donald Berwick (DB): That's a tricky question because medicine operates always under uncertainty, and so there is no on/off button here. The question is one of decisional values: When do the risks appear to be low enough that the benefits outweigh them? And when is the degree of certainty about the evidence such that we believe that we know the risks and benefits. It's a judgment call. There are very high standards for peer review publication. And a classic answer to your question is we'll know the facts to a sufficient degree of certainty when a medical treatment has been shown to be statistically, significantly better than its alternative, according to a well

done clinical trial in a mainstream peer-reviewed journal. At that point, most clinicians would say, "Well, I'll take that at the moment as sufficient evidence at least with respect to the population that was studied in the publication." Peer-reviewed literature should be one standard.

More recently, methods of meta-analysis have developed that give us the ability to look across studies and concatenate them formally into a summary assessment. One leading group fostering such systematic reviews is the Cochrane Collaboratives, which are a worldwide network of centers that essentially assemble existing studies and try to determine where the weight of evidence lies to a degree of certainty sufficient to say whether an intervention "works" or not. The usual threshold for declaring a finding as positive as you know is $P < .05$, which means that there is only one chance in 20 that we would have observed this effect by chance.

ECT: How large does the effectiveness study have to be? Does there have to be a control group? A meta analysis?

DB: There are three different questions here. The size needed to demonstrate effectiveness depends on the size of the effect one is looking for and on the variability in its manifestation, and so there is no firm answer to that. The size of the needed study depends on the underlying distribution, the statistical distribution of the phenomena being studied. So it's a very tough technical question. In a good publication, the researchers will discuss the so-called "power" of their study. Each study should report its power to detect an effect of certain size; so a good paper will say that this study had, for example, an 80% chance of identifying an effect of a specified size.

Whether there needs to be a control group is a controversial matter. There's a canonical hierarchy of forms of evidence which was first developed in Canada in the 1970s and then adopted quite widely in the global clinical scientific community. The hierarchy grades evidence. The classic grading has three levels. Level One evidence is controlled clinical trials, especially blinded, randomized ones. Level Two evidence includes non-randomized trials like epidemiologic studies, cohort studies, case control studies. Level Three evidence comprises historical control trials or case series. The general view is that the studies of the first type—randomized controlled blinded prospective trials—are the Cadillacs. They're the ones that are needed to place inference on truly solid foundations. Unfortunately, we don't have many such studies, so often we are forced to use studies of lower design quality and do our best with the data available there. Some of those studies will not have a control group. They will use the group studied over time, like a cohort study instead of a case control study.

Meta analysis studies other studies, and it's a very well-developed method but it doesn't generate new evidence. It re-analyzes

old evidence. It's a very powerful technique, very well-developed by leading statisticians around the world, mainly Oxford. And I think it's now achieved the status that most people would say a well done meta-analysis, sometimes called a "systemic review," is a pretty solid way to look at a body of evidence.

There is a deeper problem here and that's epistemological. It's certainly my belief, and that of a growing number of scholars in clinical evaluation, that the controlled trial is in fact an inappropriate assessment vehicle for some kinds of effects. Randomized controlled trials are very, very difficult to conduct, and there's an interesting body of alternative research methods now which goes under the guise of "realistic evaluation." Realistic evaluation is working with the observation that controlled trials often are very unhelpful, especially in complex systems, and it's asking the question of how, in studies without controls—observational studies and other forms of studies—we can learn as much as we could from control studies. I've been won over by that group. In fact, I've written papers on the value of other forms of evidence compared to randomized control trials, but it's a controversial area.

ECT: And why is it, do you think, that we don't have effectiveness information already—because it hasn't been a priority? Why is everybody all of a sudden demanding this information, and why now?

DB: The first question probably doesn't have one answer. Properly done, prospective trials especially are expensive. We invest a lot in biomedical research, but clinical evaluation research has been kind of the stepsister, and not really given adequate support—surely not as much as basic biomedical research has. Of the billions of dollars spent at NIH [National Institutes of Health], only a small fraction of it is invested in comparative clinical trials. Second, we have a very fragmented healthcare system, and organizing research with the discipline of good experimentation takes an organized system. When we have this fragmented system, every hospital doing its own thing; every doctor is their own entrepreneur. It is hard to assemble the data.

There was a brilliant paper a few years ago by Joseph Simone. Joe Simone is a children's cancer doctor, and he was one of the people at St. Jude's Hospital in Memphis, TN, who discovered today's cure for childhood leukemia. Indeed, it's the case that treatment of childhood cancers has seen a lot of progress over the past 30 or 40 years, whereas the situation in adult cancer treatment is not quite so bright. Some of the reasons are biologic, but Simone asked the question, "How come we have made so much more progress beating childhood cancer compared to adult cancer?" His conclusion is that there are so few children with cancer that it has been politically easy to organize centralized care. Every child with leukemia is cared for in a major medical center. Basically no non-specialist doctor, no general pediatrician, would care for such a child in their office. And so, because it's kind of economically uninteresting, the treatment of childhood cancers has been assembled in centers of excellence. And therefore, every child with cancer

is basically in a clinical trial, and we are gaining tremendous knowledge regularly about childhood cancers.

In contrast, adult cancers are economically very interesting. They are very lucrative. And so we have produced a diaspora of colleges and cancer centers and hospitals that aren't cancer centers but that do the cancer care anyway. And it's very hard to assemble these patients into common platforms for clinical evaluations. So politically, gaining knowledge systematically has been difficult. And then I suppose there is the culture of medicine, everybody stroking their chin and doing the best thing they can. It's the heroic view of the individual doctor—aggregate information is not valued perhaps as much as it should be.

Why it's changing now, I don't know. Partly we have international examples that are teaching us differently. The National Institute for Health and Clinical Excellence in the UK, for example, is widely disparaged in the U.S., but it is completely unfairly disparaged. It's a brilliant organization, precisely existing to look at all comparative information and understand what works and what doesn't. We have the Cochrane Collaborating Centers who raised the visibility of this issue; we have AHRQ [the Agency for Healthcare Research and Quality], which has strengthened research on methods on comparative effectiveness. Then of course there is the cost side. We know that there is a tremendous amount of care done that doesn't help patients, and we need to turn the lights on regarding which kind of care is helpful or unhelpful in order to save money.

ECT: Let's move to electronic medical records. What role will the national electronic record system play in the effectiveness debate?

DB: One hopes that it will be a major contributor. In theory, information collected electronically at the point of care is the best information. The care-act generates the knowledge on the basis of which the care-act can be improved. The absence of an easy ability to aggregate information from the sharp end—the actual point of care—is a big inhibitor to understanding the treatments that are used and the effects on patients. Of course, an electronic medical record system alone doesn't solve that problem if it simply automates the rather broken paper system we have now. A fragmented system with inaccurate coding and bad analysis leads to automated junk, instead of giving us the knowledge we need. But in theory, electronic records should be a very effective assist to comparative work. One of the mainstays of really good effectiveness studies are registries, which allow us to study populations of patients with multiple sclerosis or breast cancer or migraine headache. With proper registries, we can assemble virtual panels of patients and look at and learn from the experience of a population. They are easier to maintain electronically.

ECT: Once evidence-based studies are being conducted, how will we protect the privacy of those being studied—could people lose their health insurance? Obviously we have the Genetic Information Nondiscrimination Act now, but do you see privacy issues as we move forward?

Member Forum

DB: I don't want to be naïve about this. I'm sure there are potential risks to privacy. Maybe a realistic way to look at it is to acknowledge that there might be a trade off here—a classical trade off between individual privacy and public benefit. The public benefit here being knowledge about the fate of diseases and the effectiveness of treatments. I guess you can't say that there's no risk whatsoever. However, I think that if you take a step back first, it isn't IT that creates the privacy threat. That threat exists even when people collect information on paper. There's a risk of violation of privacy, and there have been egregious examples where privacy has been violated through misuse of paper records. I think the IT community and the whole community are plenty smart enough to design privacy protections that really would work and make some sense. Sometimes, however, our efforts don't make sense. The HIPAA [Health Insurance Portability and Accountability Act] regulations have way overshot in their interpretation. We are really damaging patients everyday because of HIPAA. Individuals get hurt because we've written rules that have been implemented blindly without logic, and we could repeat that mistake in the IT world as well. In terms of whether people lose their health insurance—this is an issue not of IT, this is an issue of regulation. If we want to create an insurance industry that is forbidden from doing that, which is prohibited from using this information to deny people coverage, we would be doing something very rational and something that other countries have long ago done. The concept that we have an insurance industry whose main objective is to get premiums from people who won't need care is just wrong. And I would view that as a regulatory issue, not so much as a medical records issue.

ECT: The final question in this grouping is compulsory participation in population-based studies going forward. Do you have a vision of how that would work in terms of obtaining informed consent, and persuading people to participate?

DB: I would say that if a patient—any person—is in a situation where there is randomization under way, where they might get treatment “A” or treatment “B” and that's done by roll of the dice, the ethics require disclosure and permission. I'm not an ethicist but as a person, I would say people should know when their treatment is being consciously decided based on a roll of the dice. Of course people roll the dice every day today, and that takes the form of deciding which doctor's office to drive to, because what treatment they get is going to depend on which doctor they see. In order to systematize the information, we have to use randomization differently. In population-based epidemiologic studies, we're studying, for example, the natural history of congestive heart failure, or what happens to people with migraine headache on aspirin. I think a current form for this is an opt-out which involves a default. In entering care, a person is contributing information on the basis of which care can be improved, unless they opt out. Systematic learning is the default as long as there is a legitimate and well-described option to opt out. My guess is that most people would not opt out. But maybe that's the solution.

ECT: That moves us to future formularies. Once we have evidence-based therapies established down the line, how do we enforce them?

DB: Are you asking me how I think they would be enforced or how they should be enforced?

ECT: Let's go with should.

DB: This is a stratified problem, then. Some evidence-based therapies are shown to be so effective, with such a high degree of certainty, and the downside of not using them so great, that some form of strict-enforcement “default” is in my opinion called for; that treatment will be used automatically, unless consciously stopped. A good example would be peri-operative antibiotics. We know now that giving an antibiotic for certain kinds of surgery in a particular time window has a major effect on surgical infection rates. And up until recently, every doctor would have to write an order for that antibiotic at that time on every patient, and of course they frequently forgot or they didn't know or something like that. Many progressive hospitals are now saying to the doctors, or the medical staffs are saying to themselves, “We are going make peri-operative antibiotics a default. The nurse will give the antibiotic every single time for colon surgery of this form unless the doctor says ‘No.’” That's enforcement, and it works. It produces very high reliability.

I think there are even stronger cases where you want to design a system that must perform in the way the evidence says it should. For example, we know that stocking multiple concentrations of heparin in a medical surgical unit inevitably leads to mix-ups and overdoses. Current safety standards now counsel almost a mandatory design that says that only one concentration of heparin ought to be stocked at a time. Requiring that is an enforced evidenced-based move.

I think for a broader category of “evidenced-based therapies,” enforcement is too strong an idea. The “strong suggestion” might be a better way to think about it. The reason is that individual variation does matter, and you'd really need a sensor at the front, in contact with the patient, who can say “No, not this time, not this person, this is wrong.” And when we overshoot with too strict requirements, nonsense happens. For example, we know that statins reduce heart attack risk in high cholesterol patients. There are people today in hospice care on statins because the institution wants to be in accord with the “evidence.” Well that's dumb. A person who is going to die soon doesn't need or benefit from statins. That's a waste of money and it may be toxic to them. We were enforcing prompt antibiotics for cases of ambulatory pneumonia that arrive in emergency rooms. There are a number of emergency rooms in this country that are starting antibiotics on everybody that arrives, even if they are coughing a bit, just so they get 100% on the evidence-based score. Cases like that have been written about a lot recently, and I think with appropriate skepticism. It's not a great idea to enforce things where there are important judgment calls to be made.

Intermountain Healthcare (www.intermountainhealthcare.org) does an interesting thing—they allow any doctor to violate any protocol. All they have to do is explain why they did it. And then they gather—they harvest—that information to update the protocol. That is a very interesting feedback mechanism that makes a lot of sense to me.

ECT: Could you envision attorneys spending time down the line becoming experts in arcane medical procedures, as they attempt to defend physicians who are not using the “approved” treatment?

DB: Yes, I do. I think that a lot of medical care is not as arcane as it looks. Medicine has spent a long time mystifying itself. And it would be great to have a group of critics or helpers who go to school enough to realize that it’s not so complex. And if they can’t understand it, maybe something is wrong. So actually crossing this bridge between fields is something I encourage. It’s the same for management and physicians. We built these big walls, and we’ve got to tear them down. I wouldn’t say this is the only reason why attorneys might want to become better students of clinical care itself, not just defending physicians’ coloring outside the lines, but maybe realizing that it’s a case they shouldn’t take or coaching a physician that maybe this is a time to settle with more haste and precision. So I think crossing intellectual boundaries is a really good thing.

ECT: What do you foresee happening to treatments for which there is perhaps not as compelling evidence of effectiveness, but that are widely accepted? Chiropractic care is one, some even believe psychotherapy, and then dietary supplements, other treatments that some of our friends in the government think may be fraud but other individuals and practitioners swear by.

DB: First, not in your question but needing comment, is whether there is compelling evidence of their ineffectiveness. We have trouble getting effective things to be used regularly, and we have even more trouble to get things proven ineffective to leave practice.

Here, you are asking about a middle ground where we lack evidence either way. We don’t know based on evidence whether something works or not. I’m pretty orthodox on this: I think if one has a treatment—be it chiropractic, or psychotherapy, or dietary supplement—but is unwilling to have it subjected to systematic study, then I question the basis of knowledge. I don’t think we should move into faith-based medicine. I think that’s wrong—it’s a mixing up of two worlds. So I do think that the door should be wide open for any form of potential healthcare to be subjected to systemic study. But it should also be obligated to be subjected to systematic study.

That said, there’s a soft edge to this. We certainly know more than we can prove, and there will always be circumstances in which there will be growing confidence that something’s up—something is important—and that, even if we don’t understand it yet, it appears to be a good thing. I’m sure

areas of nutrition, physical environment, for example, patient-centered care, will have that property. It will be hard to prove their effectiveness formally, maybe for a long time, but intuition and growing consensus will strongly favor them. In the long run, however, just saying, “It works because I believe it works,” isn’t going to be good enough—certainly not in a resource-constrained environment and probably not even in one that has all the resources it needs, but simply cares about safety.

ECT: How will or should evidence-based medicine account for regional variations in medical practice? For example, Cesarean section rates differ from one part of the country to another. Do you see physicians essentially bringing their procedure frequencies into alignment across the country?

DB: I hope so. The degrees of variation we see are indicators of nonsense and ignorance. As long as we remain ignorant because we lack appropriate studies, then we would certainly expect variation to be there. But that variation then ought to be used to generate knowledge about what works and what doesn’t. Right now we are living with unconscionable levels of variation which make it impossible to believe that both ends of the frequency distribution are correct. We need physicians to be curious about that difference and then help resolve our confusion by understanding what’s known and bringing their practices into more alignment with the evidence and with each other. The default—the alternative to gaining knowledge—is to say simple-mindedly that the low rate of use is best or that the high is best. That’s not very sensible either. We need an ongoing dialogue between knowledge building and reduction of practice variation.

ECT: Finally, NIH recently had a conference on the value of family history in a clinical setting. It’s certainly an historical part of our system, but the conference concluded that there wasn’t much hard evidence for its benefit. So should physicians stop using it? How do you see that playing out in the future? What’s the role of common sense in the practice of medicine?

DB: You know I wasn’t aware of this conference. It doesn’t smell right to me. This may be a case where the wrong epistemology is being used to ask the question. Family history would seem to me valuable. At least, I’d say, the burden of proof is definitely on those who say it has no value whatsoever. As we get closer to test tube tests, genetic tests that actually show us risks, of course, then family history will be a poor shadow of what we can know about a person, but I guess I would at this point say I might choose common sense. **C**

AHLA President Elizabeth Carder-Thompson can be reached at ecarder@reedsmith.com. Look for her future interviews with key policy- and decision-makers in future issues of AHLA Connections and special online-only issues at www.healthlawyers.org/connections.