

**Interview with John Ware**  
**March 24, 1998**  
**Boston, Massachusetts**  
**Conducted by Edward Berkowitz**

**Berkowitz:** Let me ask you first about yourself. Did you grow up in California?

**Ware:** My father was career military, so we lived everywhere and nowhere for any length of time. We lived in Japan and the Philippines and all over the US. I went to fifteen different schools, three different high schools, as many as three different places in one year and never anywhere longer than 2 ½ years. I think that was the record.

**Berkowitz:** What was your dad's specialty in the military?

**Ware:** He was in the Air Force and was in flight engineering. He was the highest level of enlisted man and in the Strategic Air Command and the Tactical Air Command. He was in all the wars.

**Berkowitz:** So he was in the Army Air Corps.

**Ware:** Right, in the Second World War and in the Air Force after that. He was career and retired at a fairly young age. We tended to live in California. That was our base and that's where most of our relatives were. In between assignments, coming and going, vacationing, California was where we collected.

**Berkowitz:** You went to Pepperdine, which is in L.A.

**Ware:** Yes, it was in L.A. when I went to Pepperdine. At that time they were acquiring Malibu, a beautiful campus overlooking the ocean. But they were building that as I was leaving, so I never had that advantage. I did my Bachelor's and Master's at Pepperdine, both in psychology.

**Berkowitz:** Why psychology?

**Ware:** I started in business in college, then went to math. In the second half of my freshman year or maybe the first half of my sophomore year, I took a general psychology course and got very, very interested in it. I thought my interest was in clinical psychology, but experimental psychology—although I didn't know what it was at the time—was fascinating with its laboratory experiments. It was out of the study of that that I became fascinated with measurement. How do you get a number for something that you can't see? How do you validate it? How do you study its reproducibility? Something like an attitude or a health perception, a preference or a personality variable or intelligence. I started studying psychological tests. Most of my classmates who were studying clinical viewed that as the dirty work. They had to get through those course to go into clinical practice. I always thought that the real advantage that psychologists had over psychiatrists was this rigorous training in psychological testing, but it seemed that most psychologists left that behind when they went on into clinical practice. They really gave up the real advantage that they had in a medical setting. I then had a number of different attempts at doctoral work.

**Berkowitz:** This was around 1966?

**Ware:** Right. I took a job as a clinical psychologist with the lowest level of clinical psychologists in the county, and I worked with rehab patients. We were trying to get these chronically unemployed, chronically handicapped back to work and back to functioning. That was at Rancho Los Amigos, my first job. I really didn't feel like we knew what we were doing. We didn't have any good data. Out of that, and some other work I was doing at USC in the School of Medicine, I really got interested in the patient point of view, both on the doctor-patient relationship and also with respect to their health, their functioning and their well-being, although I

didn't understand those concepts then. I didn't even have those words in mind, but in retrospect, that's what it was, what people are able to do and how they feel. Right in the middle of that job, I had an opportunity to start clinical training at UCLA. I left the job and went over there to do that and was very unhappy. Clearly my thing wasn't clinical. The research that I was doing as a part-time research assistant at USC in the medical school where we were trying to measure the doctor-patient relationship and change in patients over time, that's what I was really interested in.

**Berkowitz:** So that precluded being a doctor, which is always the issue for people like you.

**Ware:** Exactly. That was a big issue for me and almost everyone of my friends. We were struggling with whether or not to get medically trained. Some of my friends elected to do that, and I watched that with great interest. In fact, when I left USC in 1972 and went to the medical school in Southern Illinois—a new medical school that didn't have any students—I was offered this job in a department called Health Care Planning, which was *way* ahead of its time. It was basically to monitor the health care needs of the people and then integrate that into the medical school curriculum and design a training program that would serve that population and attract the graduates to stay and practice in the bottom half of the state. It was in Springfield and Carbondale. Everything else in Illinois was in the top of the state, all five medical schools.

**Berkowitz:** I know that's a really poor relation even to Northern Illinois and definitely to Urbana.

**Ware:** This was all targeting the bottom half of the state. So I went there to be in that department. One of the things that we discussed at that time was that at some point, when I finished my other work, I would go back and get medically trained. To make a long story short, I think that every time I seriously considered that, when I thought about taking myself out of my

field for four or five years to get that training versus what I could accomplish with what I was doing, it never looked like the right thing to do. I still wonder what I might have been able to do better if I were a card-carrying clinician and had that experience. Many times, even recently, I really wish I had a better understanding of clinical decision making and clinical practice, because where we are in the field right now we are trying to integrate this new information into improved medical decision making in real time. I don't know how to do that. I've never made a medical decision.

**Berkowitz:** So you have to understand the culture of the doctors. If you're not a doctor you can't really talk to them.

**Ware:** That's right. Some of the biggest barriers are these cultural barriers. So what have I done? I've worked with some of the best physicians in the country. The essence of this field is my co-disciplinary collaboration.

**Berkowitz:** It seems that they need people. Their research technique is pretty bad. What you have is big-time research technique. They're not very sophisticated about how to test hypotheses or even how to write. It's just a different thing.

**Ware:** Well, *some* of them. It's hard to say anything about them that's true of all of them, but some of them are very well trained in science. I think the thing that is most universally true of their lack of training is in measurement, even in clinical measurement which they get maybe in clinical epidemiology. They don't get the full benefit of psychometric theory and methods. That discipline, which is what I did my PhD in, has an awful lot to offer health care. I've made my career on that.

**Berkowitz:** So you're a psychometrician, which means that you measure how people think?

**Ware:** We measure achievement; we measure aptitude; we measure interest; we measure attitude. But very early on in my career while I was still in graduate school—in fact, I was just starting the graduate program that I ultimately finished, an educational measurement and statistics program at Carbondale that was relatively small and very intense faculty-student relations, a lot of tutorials and individual studies. I studied that and basically, very early on in those studies, I got my first federal grant, which was a measurement grant, in 1972 in the summer. It was one of these solicitations. To be very truthful about it, I got the award because I had the cheapest of the technically qualified proposals. The project officer of that grant, Bill Lohr, has been a friend ever since.

**Berkowitz:** That's the husband of Kathy Lohr.

**Ware:** Right.

**Berkowitz:** This was in the Center for Health Services Research?

**Ware:** Right. They have a branch called the Research Methods Branch of the National Center for Health Services Research. That branch funded the work that Jim Bush did in San Diego; the work that Marilyn Berger and Gilson did in Seattle, the Sickness Impact Profile; and my work on health perceptions and patient satisfaction; and other people that really laid the foundation for where we are today. That all came out of the little branch of the National Center for Health Services Research.

**Berkowitz:** That's interesting. That's one of the few outcomes I've heard about from that agency.

**Ware:** That program, that grant and contract and everything that followed it basically convinced me that I could take great advantage of my measurement training by just working on the

measurement problems of the health care field.

**Berkowitz:** It's rather striking, looking at your *vita*. Were you stigmatized by having had this Southern Illinois University PhD? Have you found that stigmatizing? Around here no one has one, I'll guarantee. In all of Tufts probably you're the only one. Nobody at Harvard and nobody at Mass General.

**Ware:** First of all, I got tremendous training because of the University of Chicago and Northwestern, some of the classic psychometric work that was done in the northern part of the state, such as the work that Fisk and Tyler did on developing indices of what are good measures. Tyler was a student of Fisk. Fisk was one of the gurus of what is called "construct foundation," Campbell and Fisk. They wrote the classic articles on convergent discriminative validity. One of Fisk's students, Tom Tyler, taught me my scaling course and a multivariate statistician, Bill Miller, taught me my factor analysis courses. I'm going to get to the answer to your question. By being funded, basically the whole country became my course. I would travel out to California and meet with Andrew Comrey, who is a quantitatively-oriented psychologist at UCLA. He was probably the first person in the country to ever factor analyze the MMPI [Minnesota Multiphasic Personality Inventory]. He was studying the inner structure of that personality measure and using factor analysis and other multivariate techniques, so he did some of the really important work on how you develop a conceptual model of something. His was personality. How do you develop measures of that? How do you know if the measures are working? Basically, I exported a lot of his thinking out of personality into the health care field. I would fly out and meet with him three or four times a year, so I wasn't just learning from my faculty. I was learning from some of the most important people working in the field. This was a tremendous opportunity for me.

**Berkowitz:** So what you're saying is that you do have a degree from Southern Illinois, but you happened to get into this field at a time when it was expanding.

**Ware:** Yes. I was working with people who had a lot of very practical things to teach me that have served me very well. But I can tell you, now I'm on the faculty at Harvard and I go to parties and it's typical when people say, "Oh, where did you train?" When I say it, it's almost like, "Oh, I'm sorry to hear that."

**Berkowitz:** It's not on their map. They might have heard of Urbana.

**Ware:** Yes. And when you say, "I trained in Illinois," I always go out of my way to say *Southern* Illinois. It's a wonderful campus, a wonderful university and it was perfect for me. I was working with a medical school as an assistant professor, and I was a student in the graduate school. Because they were two different schools, that wasn't a conflict. And the dean—I have to plug Dick Moy who was the first dean of the medical school—basically released me to do my research. I was there for three years and basically left as they got students. So here I was, well-funded, I was on institutional hard money, and I always did twice as much research as I was funded to do because of the release time that they gave me. So it was a wonderful opportunity for me.

The second thing that absolutely totally changed my life was a briefing at HHS.

**Berkowitz:** It was in the 1970s, so it was HEW then.

**Ware:** This would have been about '73 or '74.

**Berkowitz:** It became HHS in 1980.

**Ware:** So it was still HEW then. They brought all the measurement people, the groups that were working. I was the only one using traditional psychometric methods. Other people were

doing utility and other things. I was one of the ones that was using factor analytic methods. They invited me to this conference and we all presented our work, where we thought it was going to go and what we thought the implications were. In the audience at that meeting was someone from the RAND Corporation, which was just getting ready to launch a social experiment in health care. Families were going to be randomly assigned to different systems of care to determine what would be the cost of free care.

**Berkowitz:** The Health Insurance Experiment?

**Ware:** Right, the Health Insurance Experiment. Well, they also wanted, in case cost-sharing, co-payments, deductibles, co-insurance reduced the cost of care, then the big question was what about patient satisfaction. And what about health outcome? I was invited to consult with a team that was forming at the RAND Corporation to address these outcomes. It just so happened that the talk I gave on how you measure patient satisfaction and how you measure overall health perceptions—"What is this thing called self-evaluated health?"—that was the part that I had been funded to work on. So I started consulting with them. In 1974, the Secretary of Health decided to expand that experiment from a pilot in one site, that being Ohio.

**Berkowitz:** Was that the Assistant Secretary of Health?

**Ware:** I was told that it was a Secretary-level decision. It was the last thing that was done in, was it the Ford administration?

**Berkowitz:** David Matthews?

**Ware:** Would that be '74?

**Berkowitz:** Ford was president 'til 1977, so it would have been the last Nixon one maybe.

**Ware:** Anyway, the decision was made to expand the study to six sites and the measurement not



just of costs but of these outcomes. It was very clear to me, and one of my colleagues at the time, Bob Brook, suggested, “Why don’t you just do your measurement research at RAND because we’re going to be doing this for a long time?” But I really went to RAND not to stay there. I went there to provide a technical function, to develop questionnaires for this experiment, and everyone made it very clear that when I was done with that I should move on. I didn’t really expect to stay.

**Berkowitz:** This was in 1975 that you went to RAND?

**Ware:** Right. I started consulting in 1974 and then moved there in the early months of ‘75 and stayed there fourteen years. That was the other perfect environment: a very large, well-funded study; had big appetite for measurement; and the principal investigator, Joe Newhouse, really was committed to the non-economic part of this study. So as long as we were making methodological advances—and we certainly did. That study led to a number of breakthroughs that we still are benefitting from in our understanding of health and satisfaction from the patient point of view, *we* how *me* measure it. It changed the thinking a lot, for example, the use of self-administered questionnaires. Unthinkable before then. The government’s still doing mainly all interviews, but we based everything on a self-administered questionnaire booklet. Everyone said, “You can’t do that, and if you do it, people won’t fill it out. You won’t get high completion rates. The people in the Southeast won’t understand it,” and on and on and on. We challenged all those preconceived notions, used self-administered questionnaires for five years, and I don’t think we ever got below a 90-95% response rate. The data quality was fantastic. We also learned that that was much more data than we needed. That was what we call now “psychometric overkill.” We had hundreds and hundreds of questions in each booklet. We learned later—something we would

never know any other way—that we didn't need all that. The short form measurement era came after that. SF 36 is an example.

**Berkowitz:** That stands for short form?

**Ware:** Yes.

**Berkowitz:** So that's an attempt to compress this. I see. I'm curious about this. My experience with how hard it is to measure things is in the field of disability, like your Ranchos Amigos days, where they have to decide who is disabled and who is not. It turns out it's very hard to do. Seems to me that's an intractable problem, that you're not going to be able to measure that effectively. Are there limits to this? There's no magic bullet there. It's not clear who's disabled, who's not disabled, who deserves the money, who doesn't. Maybe you could do something with replicability from person to person, but that's about it, it seems to me. Is this a field that can address things like that?

**Ware:** Yes. And I wouldn't accept those premises. I would challenge them. They are empirical. They can all be addressed empirically. When we designed the Medical Outcomes Study, which was the study that followed the Health Insurance Experiment—a true randomized trial that was huge, long, had a budget of eighty-five million dollars in 1980 dollars, but it basically studied the well end of the spectrum. People were younger and well. And it studied the non-aged spectrum. No one in the study would be over 65, because it was unthinkable to use either cost-sharing or prepaid care in Medicare when that study was being started, so we didn't even include the elderly. It was just unthinkable then and we didn't even include them. So the Medical Outcomes study was a non-experimental study, much smaller, but everyone was chronically ill. We started thinking about this study in the late '70s, and we started designing it in

the early '80s.

**Berkowitz:** Was that also a RAND study, the Medical Outcomes Study?

**Ware:** The Medical Outcomes Study actually began at the University of Chicago. Alvin Tarloff was the retired Chairman of Medicine and was well-known for his GEM\_\_\_\_\_ [Graduate Education in Medicine\_\_\_\_\_] work, his controversial studies of graduate medical education. The GEM\_\_\_\_\_ study led to a big debate: how many specialists do we need; how many generalists do we need? It was very clear that when you get them all in the room there were big arguments about whether the specialists treat sicker patients or more complicated patients than generalists. There's a lot of evidence against that, that in fact probably a third of the practice of a specialist was patients that didn't need to be there, but that's how they kept their offices full. There was too much overlap in the severity distribution, and there were big arguments about who was the most effective. And those are still going on. So the Medical Outcomes Study, his interest in the Medical Outcomes Study, the only way to answer this question, "Who should treat the chronically ill?" was to do a Medical Outcomes Study. So the Medical Outcomes Study was originally thought of as a manpower study. "Are family practitioners or internists or the modal specialists—like a cardiologist for heart disease, or a diabetologist/endocrinologist for diabetes—most cost-effective in treating the chronically ill. We began to plan the Medical Outcomes Study to address that issue. That study, because it was going to begin in office practices, faced some practical constraints that we really didn't have in the Health Insurance Experiment. Those practical constraints led us to develop shorter measures, measures that we could administer in the clinic. We actually sampled patients who were in a clinic and we screened them and wanted to assess them. We needed a short form to do that. The SF 20, the first widely

used short form that we created with 20 questions, actually came from the Health Insurance Experiment. It was a very short battery, sub-set of items. So the first measure that we used in the Medical Outcomes Study was really a short form developed from the questionnaires used in the Health Insurance Experiment.

**Berkowitz:** What would be a question on that short form?

**Ware:** How would you rate your health? Five choices, excellent to poor, would be an example. How limited are you?

**Berkowitz:** ADL kind of questions?

**Ware:** Much higher. ADL would be down in self-care. We had one of those questions. Whereas ADL and IADL maybe would have 80, we only put one. The reason is there's hardly anyone down that low on the scale. We focused above that where the great majority of the population, even the chronically ill population, scores. We would ask questions like, "Are you limited a little, not, a lot, or not at all in doing your everyday physical activities like walking and climbing stairs and moving furniture around in the house?"

**Berkowitz:** That was the SF 20?

**Ware:** That was the SF 20. Let me go back to a question you asked me earlier and I didn't answer it. The Medical Outcomes Study started at the University of Chicago. I was brought into it, before it started, by Al Tarloff and Ed Perrin. Ed Perrin, the former head of the National Center for Health Statistics, was advising Al Tarloff on the design of the study, of the statistical issues. But this was a study that was going to largely use patient-based measures. Ed said talk to John Ware. So one of the nicest things Ed ever did was to get us together. We started talking in the early 1980s and began to conceptualize this study, which was broadened in its focus from just

what provider groups should treat these chronically ill to a study of in what settings should they be treated? Is there an implication? Can chronically ill people have equally good outcomes if they're treated in a prepaid plan? A prepaid group practice HMO as opposed to a fee-for-service plan. That study kept competing between those two objectives. Ultimately the system comparison, the prepaid health care versus fee-for-service care comparison—was the one that rose to the top. When we had to embed one within the other, that was the one we gave the highest priority to, although we'd done some very useful comparisons of specialty care versus generalist care, that was embedded in the system analysis. Our first sponsor was the Robert Wood Johnson Foundation. They let a grant to the University of Chicago. Clearly, at that point, the Medical Outcomes Study was going to be a collaborative study. At that point the collaborators included me at the RAND Corporation; the principal investigator, Tarloff, at the University of Chicago; Ed Perrin, the statistician at the University of Washington; and two other health services researchers, Mike Zubkoff, an economist, and Gene Nelson at Dartmouth. They had actually independent of us proposed their own little study of family practice and how does family practice do with the chronically ill. We said let's fold all this into one. The study ended up at the RAND Corporation because, in the meantime, Al Tarloff became the president of the Henry J. Kaiser Family Foundation. That was great from the point of view that we got the funding we needed to match the Johnson Foundation, and the Pew Charitable Trusts also put up a million dollars, but we lost our principal investigator. So I stepped up into that role. For that reason, and because we basically were using the brain trust at RAND to design the study and to plan and do the data collection, the principal activities for those early years were done almost entirely at the RAND Corporation.

**Berkowitz:** What years was this Medical Outcomes Study?

**Ware:** We began the preliminary work in 1983 and 1984, testing our methods. We actually went to the field in 1986. To get back to a point I was making earlier, I think our sponsors really very much expected us to take the tools off the shelf that we had used successfully up to that point. Me having an equally large objective for the study to take measurement to its next logical level, I wouldn't have any of that. So we spent a small fortune doing methodological work during the several years before we went to the field—millions—developing shorter tools, adding measures of other things we thought were important, developing some disease-specific measures. The short form development went on during the course of the study. It wasn't until five years later that we developed the SF 36. Actually, the SF 36 was developed after I left the RAND Corporation. Let me tell you what was going on in 1988. A guy named Paul Elwood—my sponsors called me and said, “We want you to talk to this guy”—was wanting to develop a concept that he had of outcomes management. In other words, let's manage care and let's evaluate the success of competition in health care in terms of costs and outcomes. We know costs; we need outcomes. So he was going to write a very important editorial called the Shattuck Lecture in the *New England Journal of Medicine* in 1988, and he wanted to propose in that lecture that the SF 20 be adopted worldwide as basically the tool for monitoring outcomes from the patient point of view. He called it the technology of patient experience, a very clever idea. And I said, “That's a great idea, but that's the wrong form. We don't even use that form any more.” We had stopped using it years earlier. We were developing a new short form, that was the SF 36. We talked a lot about should the SF 36 be the SF 44 or 40 or 32. He was pushing me to stay close to twenty and I was trying to add the additional items that I knew had to be in there

for that thing to work. That was the push and the pull. I compromised at 36. All that was done in the six to twelve month period after I left the RAND Corporation in June of 1988. The Medical Outcomes Study work continued at RAND; I continued the Medical Outcomes Study work here; we had collaborators at Dartmouth and elsewhere. Will Manning, an economist who did a lot of the work on the Health Insurance Experiment, did some very important economic work in the Medical Outcomes Study from the University of Michigan and the University of Wisconsin. So we were scattered all over. I think there've now been 100, at least 150 articles that have come out of that Medical Outcomes Study. Most important from a methodological point of view, it has really been the laboratory for the advancement of measurement. The use of generic measures in different diseases—people resisted that—turned out to be very successful. The use of short forms, people resisted that. It's turned out to be very successful. The use of the same short forms in clinical research that we use to monitor populations has been tremendously useful, but it was fiercely resisted during those years. You had to use completely different measures for different people and for different purposes. There was no standardization, and we *still* don't have enough standardization in this country. We have no unified approach to measuring the benefit that we produce in the health care system in this country. All the different agencies look at function—you mentioned disability earlier—limitations, health, well being. They all measure just a little bit differently. We now know that a standard measurement system, kind of a standard generic core to which we would add additional measurement for specific purposes, should allow us to link the Census to the National Health Interview Survey to the entire research program of the NIH to everything that AHCPR [Agency for Health Care Policy and Research] is doing. That was the organization that was created by law in 1989 to monitor outcomes and to do

other things. That law said, “By outcome we mean functioning—what people are able to do, well being—how they feel, and satisfaction with care.” Those three concepts were written into law. That was a tremendously important event and it gave legitimacy to the patient-based measurement movement. We took that and we ran with it. We didn’t get very much funding from the federal government, but that basically wrote the concepts into law. But we didn’t have any tools that were practical enough to be written into law, so that’s the problem we’ve all been working on since then.

**Berkowitz:** Let me just take you back. In the Health Insurance Experiment is there a Washington story you tell, or is your work really based on this methodological stuff? Do you tell a story about, for instance, one story might be that when you have to pay something, there’s a co-pay, the use of the service is less? Were you involved in any of those kinds questions?

**Ware:** Oh, absolutely. I wasn’t supposed to be, but I got hooked on the policy issues. My being hooked on the policy issues reached intensity in the analytic period. We ended data collection for the Health Insurance Experiment in 1981. I was given the primary analysis responsibility for the health status outcomes in the experiment. Another colleague, Alison Davies [my wife], did patients \_\_\_\_\_. Bob Brook did process of care and other variables. I was comparing health outcomes as a function of cost sharing and fee-for-service. That was the 1983 *New England Journal* article—Brook, Ware, Rogers *et al.* And then a 1986 *Lancet* article by Brook, Rogers, *et al.* Those two papers were the principal health outcome papers for the two major parts of the Health Insurance Experiment. One part looked at the effect of cost sharing, which we had previously shown greatly reduced the costs of care. Cost sharing reduced the costs of care 20-40%. The question was, at what price to health? Our paper in the *New England*



*Journal* said basically, “There’s no improvement in health with free care. The 40% increase in consumption costs more, but there’s measurable benefit to the average person.” To quote something that Joe Newhouse wrote later in his wonderful book called *Free for All*, a Harvard University Press book which summarizes the entire Health Insurance Experiment, he documents the increase in cost sharing and indemnity insurance plans in the country increased dramatically during the years immediately after our results came out. Basically it said, cost sharing reduced costs and its free with respect to harm to health for the great majority of the population. By his estimates, the country probably saved more than the entire cost of the study in the first few months that they started increasing cost sharing and indemnity plans all around the country.

**Berkowitz:** Yes, but *post hoc, propter hoc*. It’s a fallacy, of course. They didn’t do that because of your experiment. They did it because they were trying to save money, right?

**Ware:** But I think the experiment gave them the data when people argued against this saying, “You’re going to hurt people.” However, the thing that really hooked me on all of this was that there was some evidence that the answer to the question was not the same for everybody in the population. The reason I got interested in the policy debate was because, frankly, I was dissatisfied with the extent to which we were looking at variations in outcomes, particularly the poor and the sickest groups in the experiment. We saw, and we reported in both of those two articles I just mentioned, that, yes, cost sharing reduced expenditures substantially, the outcomes are the same on average, but the poor sick look like they’re getting hurt. That really caught my attention. They had worse outcomes in the fee-for-service part. In other words, when we made people pay a larger part of the bill, they had worse outcomes relative to free care. When we compared the HMO, the prepaid group practice HMO, with fee-for-service care, in all of those

comparisons the poor sick consistently had worse outcomes. We saw it in clinical outcomes like blood pressure control. We saw it in symptoms and disability days. We saw it in health perceptions and the other measures that I was studying the validity of. I'll never forget the briefing to the RAND board. Harold Brown, the former Secretary of Defense, was on the board. He read a newspaper during the entire briefing. And we presented a summary of results, before they were published, showing that the poor evaluated their health less favorably after five years with cost containment relative to poor who didn't have cost containment. I'll never forget it. He put down his newspaper and said, "How do you know they weren't just saying that? How do you know they weren't just *saying* that their health was worse?" The answer, of course, is that we had validated those self report measures and we knew when people said their health was worse, their health on average *was* worse. In fact, we showed in one study that the death rate was twenty-five times higher within five years for those who rated their health unfavorably at the beginning of the five years, relative to those who rated it favorably. So our answer to Secretary Brown, not in the meeting but subsequently when we thought about, "What do we wish we had said," is that at least they're being true to their word. They're dying at a much higher rate than when they don't say they're worse. That really impressed me with the importance of measurement validation when you apply these tools in a policy study, because anybody who doesn't like your results is going to attack your measures. I had learned that; let me go back ten years. In 1966 I was a research assistant in the Department of Psychiatry at Los Angeles County General Hospital in USC Medical School. We got a new governor, and the governor said to the hospital, "Prove to me that we should continue our in-patient mental health program in the state." One of my first tasks was to go and get all this information about how many admissions, how

many diagnoses, how many beds, how many ECTs we were doing, how many drugs, and he threw all that out. He said, “That’s a numbers game. You’re just telling me what you do. I’ll give you that you’re spending all this money on mental health services. Show me your outcomes.” And they didn’t have any. And he *completely* dismantled the state mental health program. Maybe there were some good things that came out of that. I’ll never forget that. I was the lowest person in the department. Here I was trying to measure outcomes from the patient point of view and I said, “I don’t ever want to be in that position where I don’t <sup>doubt</sup> my outcomes,” because you are so vulnerable. That made an impression on me in the back of my mind. Later when we went into the health care cost containment debate, I said that if we were going to reduce the cost of care, provide less care, I want to know the outcomes, because I know the way the debate is going to go if we have no evidence to the contrary.

**Berkowitz:** That’s a very upbeat view of the policy process. There are lots of counter examples I could give you where there’s lots of data but it doesn’t make any difference. People believe the same thing because that’s what they want to believe. For example, they believe in workfare for welfare beneficiaries, but there’s not very good evidence about the outcome there. But they believe in that because that’s what they believe in, so it doesn’t matter. You could show them all sorts of things.

**Ware:** Oh, I agree. I think something that proves your point is if you read the transcript when the Health Insurance Experiment data was finally discussed in the Senate. Finally a senator did a sixty second sound bite; he got it exactly backwards. Here’s an eighty million dollar experiment that, when it finally had its moment being debated in testimony, he got it exactly backwards.

**Berkowitz:** Historically, there were several experiments in that period—taxes, health, they were

the two biggest, but there were others—that had no effect on the debate. People used the data in ways that they wanted to, selectively, to show whatever they wanted to believe. That really is interesting then, isn't it? Then the Reagan administration came along in 1980 and said, "We're not doing these big data collections because they're nonsense and we know what we want anyway. It doesn't really matter. We know what we believe. What difference does it make what the data is." It's a religion. It's an interesting policy problem. Here you are developing this measurement technique for which, ostensibly, there has to be some reason. I guess maybe the answer for you is maybe not policy but clinical practice in the hospital. There is interest in that, right, tremendous interest in your work within the hospital in more of a private way than in policy. Does that make any sense?

**Ware:** Where we're moving now is to not just use these tools in these large social experiments and for research purposes, but to really use them in management—managing of populations of patients and, the real challenge now, the individual patient.

**Berkowitz:** Right. It seems like the world has come to you. Really this is managed care's heyday. Probably people really do want to know.

**Ware:** Oh, absolutely. And this is having a tremendous impact on the way we measure. You mentioned when Reagan became president. Well, when Reagan became president, I was in the social science department at RAND. After he was president, I was in the behavioral sciences department at RAND, the political sciences department. It was the same department. We changed our name, because what we were doing was not fundable. We did our work under a different label, because he really had no interest in this hodgepodge of social science work. The real challenge now is to take measurement and make health care outcomes and patient satisfaction

part of the health care data base, and use that to improve the cost effectiveness of care. What we're working on today is what I call a Unified Measurement Strategy: how will we link all the data bases. The population data base defines the norm, basically tells me what we should be trying to do for you. It should be trying to make you equivalent to a healthy, productive version of yourself without the burden of whatever problems you have. The goal of treatment is to make you your best. The norm becomes that benchmark. Then when we evaluate the treatments like in clinical trials—what works, what drugs, what surgeries—we want to measure outcome the same way that we do in the general population, so that we can say here's what a normal well person looks like. Now we know that there are huge variations in outcomes even when virtually identical people get virtually identical treatments. The variation in outcome is at least as large as the variation in practice styles. We want a measurement system that we can use one patient at a time, so I can ask myself if it's worth treating this patient with this treatment, and am I getting the result for this patient that I should be getting, what the clinical trial said I should be getting? How do I decide who are the third of patients that are going to benefit from this treatment a lot; who are the third that are going to be in the middle; and who are the third that are going to be no different than placebo? Basically that's the information system we're trying to develop right now. That's where measurement development, in my field, is right now: how can we do that in a very practical and precise way? What that demand has done is send us back to the drawing board, back to psychometric theory and methods and a whole, fundamentally different approach than the one we've used up until now, because the one we've used up until now will never give us the precision that we need to monitor an individual patient in a practical way. So the short forms that we've constructed by taking a subset of questions that represent the measurement scale—we

know now that that short form is doomed to fail in measuring an individual patient. If I were going to pick thirty-six questions to measure a patient, the questions I would pick would depend on the patient. What kind of morbidity does this patient have? Is she high on the scale or low on the scale? So what we want to do is basically individualize the selection of questions according to what your health problems are and where you are on the scale. This leads us to a completely different setup.

**Berkowitz:** More like replicating the clinical experience that the doctor has. That's what a doctor would do too.

**Ware:** It's analogous to that.

**Berkowitz:** He's going to say, "How are you, Mr. Jones?"

**Ware:** Right. And from there she'll go highly individualized.

**Berkowitz:** Bringing the theory in from the back door. "Can you lift your hand?"

**Ware:** Right. And if you can do all those things I'm not going to push that.

**Berkowitz:** And the doctor knows it's cloudy that day, and the pressure outside might affect the...etc.

**Ware:** Yes, so what we're doing is we're going to what has been used like in achievement testing where they do the GRE or the licensing exam or the Navy does its aptitude test. Nobody fills out all the batteries any more. Nobody does them in paper and pencil. They're done by computer, and it's called computer-adapted testing. If I ask you a question and you get it right, I ask you a harder question. If you get that right, I ask you a harder question. If you get it wrong, I ask you an easier one. I only ask you the questions that measure where you are. The short form is a highly selective process from an item pool, which is all the questions that measure all the

levels in whatever concept you're trying to measure. The computer, after every one of your answers, says what is the information value of every question in the item pool and what question is going to give me the most information about where you are?

**Berkowitz:** Like in the old fashioned days where they did interviews, they would have, "If yes, to Question 36..." and the computer can really work through that. That would take pages and pages.

**Ware:** Right, but it's much more than a skip pattern. It is literally computing a score and is picking a question on the basis of an estimate of your score.

**Berkowitz:** That's interesting, the dynamic element there.

**Ware:** Right. But in order to do this, we had to have a model of health, we had to have all the experience with all of these items, we needed all these long forms. So we're developing these systems now, and we're beginning to test them. They are going to greatly increase precision, and they're going to cut respondent burden by at least two thirds. *And* they're going to be incredibly cheap. We're also going to take advantage of the new technology in computing and voice recognition, telephone and television. These things can be run on any PC twenty-four hours a day over the telephone. The computer can simulate the voice or use a human voice. People can respond by voice or they can use a touch tone pad or whatever. We can do assessments very quickly, very cheaply, but meeting the clinical standard of precision.

**Berkowitz:** Is that something a hospital, a health provider, would buy and then call in to a central place? Is that how it works?

**Ware:** Right. Or you could load the software onto your own computer.

**Berkowitz:** You say you could do these things, but I know that people are not very bright at the

actual operational level. In fact, it's really very hard to get people to know how to do all this.

**Ware:** It's a huge overhead expense initially, there's training.

**Berkowitz:** What about the level of competence? I see the report on your desk about "The Computer-Based Patient Record." I read that. That's really very upbeat about how you could integrate all these data bases, but in reality there's a guy in charge who scribbles his comments down. That's what happens over there in the hospital right now. With all this technology, you have to be aware of people's ability to implement that. Their VCR has tremendous capabilities, but they can't program it. That simplification must be a very difficult problem.

**Ware:** Right. And it differs from country to country. We're now working in forty-five different countries. The notion of the standardized form and filling it out is much more American than it is western European. But even if you use an interviewer, you could put an interviewer on the phone or face to face with a laptop between the patient and the computer. This would tremendously cut the cost of an interviewer-based health assessment. In addition to this very important practical reason for doing this new methodology, the other reason is so that we will get a confidence interval around each patient's score. What I need to know is, "OK, this is your score. Is that score better or worse than it should be? Is it different from what it was before? Are we maintaining your health, are you declining, are you getting better?" For me to use this data at an individual level, I need a confidence interval. That's the real advantage of this system. It gives me a person-specific confidence interval around the score. Right now, we estimate that confidence interval to be the same at all levels of the scale and for all people, because we base it on the old psychometric method where we took a reliability coefficient which tells us how noisy the measure is and we took the population standard deviation of measurement which tells us how



much variability there is and applied that to everybody. That means that if I am inconsistent in filling out the questionnaire, your confidence in it was right. If you're in a population of people who are very different, standard deviation is large. We all have larger confidence intervals. But if I take that same questionnaire and put it in a pile of people who are very similar, the confidence intervals narrow. That doesn't make any sense at all. Your confidence interval should be based on your consistency. We now estimate those independent at every level of the scale, and we estimate it individually. We know now that ninety percent of the people are very consistent; I can estimate their score very quickly. But there's another five to ten percent that aren't, and I need to ask them more questions.

**Berkowitz:** That's interesting. My naive response to that is I fill out the consumer satisfaction surveys—like when you buy something like a warranty or stay in a hotel—what I notice is that if you're in the neighborhood of something, you stay in that neighborhood. If you put excellent for one, you're going to look at excellent regardless of the next question. Could be a totally different question. Maybe that's a common phenomenon, but I notice I do that a lot. You're in a norm of some sort; when you're in excellent you're thinking excellent; when you're in poor you're thinking poor. It takes a lot to get you over to the other side. Therefore the reliability from question to question is not so clear, at least with me and probably with others. I imagine that's not uncommon.

**Ware:** We hope that what you said isn't true. We would like each question to discriminate the feature of care or services that it's supposed to. We know that people have a halo effect, that they have a general attitude: I'm really very happy and I give happy responses.

**Berkowitz:** I don't often understand the question when I'm answering such questions. I don't

get the question. It looks like the other question. It's not exactly clear to me what the question is, so therefore my outputs are similar to my previous one. Maybe that's a problem with a badly designed survey, but definitely I do that.

**Ware:** Well, if you don't understand the distinction the question is making, the survey is a failure.

**Berkowitz:** But these are really smart guys that are doing these marketing surveys, supposedly.

Anyway, let me ask you two more questions.

One thing we didn't make clear is your transition from RAND to here. That's a big deal, across the country. What was the draw here?

**Ware:** Good question. Actually, there were a lot of us that left within a few years. RAND is probably one of the best full-time research organizations that one could possibly work at for many reasons. But I think it's a much better place earlier in your career than it is later. Later in your career you cost the project you work on more, so people are less likely to put you in a small amount in their project. RAND doesn't really have a large endowment. It goes from contract to contract. For senior people, as you get more and more senior you're doing more in your field that really isn't chargeable to a particular project. That's when the model that works so well in the first five or ten years of someone's career—you could hardly find a better place to be but after that point...

**Berkowitz:** You're doing precisely the things an academic would be rewarded for, but you're not able to build them in grants.

**Ware:** That's right. They're not chargeable.

**Berkowitz:** Might not be a in place where they give you a lot of credit.

**Ware:** That was one consideration. The second consideration for me was that I was entering a new phase of my work. My methods were now the end product of my research. Basically, I was developing tools and the obvious thing that needed to happen to those tools was that they needed to be used in the real world, not just in the next big policy study or the next big experiment. We had no patients at RAND, we weren't delivering any care. RAND didn't really care about these tools except as a vehicle in support of research. I wanted to begin to standardize these. I wanted to define them. I wanted to protect them. I wanted to use them in a variety of different settings. I really thought I should be in a health care delivery setting. I considered the UCLA megaplex. I looked at Boston; Boston was a real medical mecca. But the other variable that I was considering at the time was that I had gone to a briefing on the east coast and someone told me that industry was putting more money into medical research, actually several times more money than the federal government. I was thinking about that and thinking, "Who really cares about outcomes the most?" Well, it's the people that have products that produce outcomes. They're on the other side of the table from cost containment. Who are these people? They're drug and device manufacturers. So I was beginning to talk to them. I spent the last few years that I was at RAND talking to the front office about what I called the new sponsorship structure. These were the people that really wanted to invest in the measurement of health care benefits, because they wanted to be able to prove their benefits. Actually they wanted to be able to prove outcome in the terms that mattered most to the public. This wasn't really on the horizon at that point; that wasn't the way people were thinking about this. So all of these things led me to be in a different kind of an environment, an environment where patient-based assessments would actually be used not just for research purposes where your interest stops when the project is over. If we were

going to weave this into the data base, I needed to be closer to that data base. I wanted to be working with different specialists that were treating orthopedic problems or cardiovascular problems or whatever. I wanted to be finding out what I had to tell them about these tools to make them useful, how were going to collect the data, how were we going to process it. The obvious follow-up from the Medical Outcomes Study was to make these tools even easier to use and more practical, and to develop information that would tell clinicians what the numbers mean and how to interpret them. I wanted to be a in a health care delivery setting. In retrospect, this didn't turn out to be the best one.

**Berkowitz:** You have a lot of places to go out for lunch. So it's nice neighborhood to be in. The best neighborhood probably.

**Ware:** The move was out of the research ivory tower into the health care delivery setting that wanted to measure these things as part of its own survival, its own quality-improvement effort. Again, that was way ahead of its time. That still was an uphill climb. But I think that we really launched the practical measurement era: make it practical, get it into practice in the years afterward. These wonderfully elegant long form measures were made more practical, easier to use, easier to understand. *And* my research went from eighty percent federal or foundation to eighty percent industry sponsored, twenty percent federal or foundations.

**Berkowitz:** I was thinking while you were saying that, there was a time when the work that these companies that made drugs weren't interested in outcome; they were interested in getting the drug approved and put on the market. Now it's a little bit different. You're talking about a more sophisticated outcome, which I don't think the Squibb Company would have cared about a couple of years ago.

**Ware:** Squibb—I think it was Squibb—funded a classic study in 1986 in the *New England Journal of Medicine*. There were three different drugs for treating hypertension, all were equally safe, all were equally efficacious in terms of blood pressure control, but one of them had a much better quality-of-life profile than the others. That was a reason not to use the other drug. That article, Krug and all, in the *New England Journal*, 1986, and the editorial that accompanied it changed forever the way we evaluate drugs.

**Berkowitz:** It turned out to be a Squibb drug that had the better quality of life, not coincidental.

**Ware:** Their market share went up. Their stock value went up.

**Berkowitz:** They learned to speak to the physicians in forms that they could really use it.

**Ware:** And to the public. “I want you to control my blood pressure but I don’t want to be tired and impotent as a result of that. If that has to be a consequence, I’ll make the choice, but if there’s a certain drug over here, even if it will cost a little more, I want somebody to pay that because I want safety, efficacy and energy and libido.” Then the employers got very interested. When the employers discovered that untreated medical conditions were costing them more in indirect costs, much more, than they were saving with reduced access to health care, I think that is what really fueled the outcomes movement, the accountability movement: holding providers accountable for their effects on what people were able to do and how they feel, which is what we call health-related quality of life. That whole thing really got its momentum going when employers realized that there were two costs that they were dealing with here, and labor cost was a much bigger cost than health care cost. Health care cost was trivial. Even a small percentage change in labor costs dwarfed your health care costs, so they really wanted to understand.

**Berkowitz:** That’s why they got involved in disability management. They wanted to reduce

their Workers' Comp costs and their long-term care costs. But, of course, that contradicts what you were saying earlier in a way, doesn't it? You were saying that this cost sharing stuff was actually health-neutral. But you're saying it's really not.

**Ware:** It was neutral on average, but what we really learned from a number of studies, and I don't think there's anyone left who doesn't now accept the fact, is that it was very naive to think that the average was the full story for the population. And second, to generalize from the average cost and the average outcome to subgroups in the population, particularly the most vulnerable—the publically insured under Medicare and Medicaid, the elderly, the chronically ill and those who are at the intersection of all those variables—to assume that worked great for middle and upper America—middle and upper in terms of social class and working and well—we'll now apply to the poor, sick, elderly, less-educated and we're going to get the same result.

**Berkowitz:** Right. A little bit like the differences between average cost and marginal cost.

**Ware:** Yes.

**Berkowitz:** OK. One other thing I need to ask you about: I see that you are on the Board of Directors for the Association for Health Services Research. That's one of the things that I'm supposed to find out about, this disciplinary world that you're part of. Would you identify yourself as in the field of health services research?

**Ware:** Right. Very definitely.

**Berkowitz:** Even though you're also a psychometrician? Primary, secondary?

**Ware:** woul I've virtually done nothing outside of the field of health services research. I certainly came after that field started and probably it was at its peak in the mid to late '60s.

Certainly the National Center for Health Services Research was at sixty or seventy million dollars then in 1960 dollars. Then it dropped to twenty million in 1980 dollars. So it was really at its peak then and went down, then came back, and it's down again. I remember I was at one of the first discussions in the living room of Bob Brook's house in Pacific Palisades when everyone was in the room talking about creating an association that would be a professional group and maybe a lobbying organization, and that we might be able to get a hundred or two hundred million dollars federal dollars to fund health services research. We might get a larger portion of the NIH budget devoted to health services research. Everyone thought we were nuts, but that's largely what happened. I did two maybe four-year terms on the board of HSR, can't remember exactly which chairs.

**Berkowitz:** This says 1989-1995.

**Ware:** Yes. That d have been two three-year terms.

**Berkowitz:** How did that happen? You just knew the folks?

**Ware:** I was elected. It was a formal election.

**Berkowitz:** How did you get nominated? You must have known somebody or done something?

**Ware:** We all came out of RAND with some name recognition from these highly visible studies. I would gather that as outcomes were becoming more and more in vogue, that probably led to my being nominated. Anyway, I got nominated and got elected and then got re-elected. That's pretty much the limit of the term. That was a very interesting period, because I was on the board during the health care reform period, the transition from Bush to Clinton. Many of the people that were on the board at the time were right in the middle of all of that. There were a lot of discussions and some of them quite strong arguments. It was clear the Clinton administration was

going to bash the pharmaceutical industry, and I felt that their “drug money,” which everyone thought was bad, to make them the fall guy wasn’t supported by the data, and it wasn’t a good strategy. They could have been a great partner of the associations. Now it’s not at all uncommon to see various lectures or workshops or keynotes at the annual meeting of a conference supported by the pharmaceutical industry. Of course, they have their vested interests, but who doesn’t? The conflicts within our organization were at least as big and biased as the ones on the outside. That was a very interesting time. I got to testify a number of times to the Clinton-Magaziner task force, to the Senate.

**Berkowitz:** On what issues?

**Ware:** Issues of how we define quality. They really wanted to bring some rigor in. That was a big part of their whole thrust that they were going to define quality and manage for quality. But the main reason I went was to deliver the message that you’re not going to save the money in managed care in the other half of the population that you saved in the first half; they’re different populations. And second, you can’t take an approach to organizing and financing care that worked with the middle and upper income groups and impose it on the sick, poor and elderly and expect to get the same result. That is one of the fundamental, biggest mistakes that has been made in health care, and it in part accounts for why Oxford plummeted fifty percent on the stock market this past year, why we’ve seen a lot of what we’ve seen. And that is that the cost containment success that was observed with the low-lying fruit in the first half of the population, to project that to the other half we were all saying was a terrible mistake. We suffer for not having good risk adjustment in our forecasting models. That’s one of the biggest uses of these new patient-based measures, to better estimate the demands for health care. Risk management is



going to be the first big wide-spread use of health status measures. The second big use is going to be to monitor and approve outcomes, monitoring the health stock of the populations that are coming into the system. I was in a meeting recently where someone said, “Our budget for this group practice is twelve million dollars and three million of that is 384 people.” That’s a real live example. They want to know if you are one of those 384 people, and when I put you under the utilization review mechanism, I want to make sure that I’m not hurting you. So there’s tremendous interest in better data for risk management, and the next step is outcomes management where we really make sure that we’re getting the benefit for you that would justify that expenditure. And if we’re not, we should try something else, which means I need to make a measure of you at the individual patient level. This is what we’ve been doing in psychological testing for a hundred years. All of our testing has been at the individual level. Whereas in health care we measure the population level with relatively crude tools. They do not have the precision that we need at the individual level. We know how to construct those, but we never have had a reason to do so up until now; now we do.

**Berkowitz:** That’s good. That unifies things. Thank you very much.