Development of Clinical Practice Guidelines
Alfred Berg, M.D., M.P.H.

DR. FERREIRA-GONZALEZ: So what we're going to do is start introducing Dr. Berg, even though he's not here.

Dr. Alfred Berg is here to discuss the development of clinical practice guidelines. He has been very active on several expert panels in this area. He was chair of the CDC Sexually Transmitted Diseases Treatment Guidelines Panel and member of the AMA-CDC panel producing Guidelines for Adolescent Preventive Services, and a member of the IOM Immunization Safety Review Committee. He currently chairs the CDC EGAPP Working Group, as well as an IOM panel examining evidence on the treatment of post-traumatic stress disorder.

Good afternoon, Dr. Berg. Welcome.

DR. BERG: Thank you. Should I just begin?

DR. FERREIRA-GONZALEZ: Yes. We have already done your presentation while we were waiting for the connection. If you would please go ahead.

DR. BERG: Thank you. Well, it's a pleasure to be with you this afternoon. I wanted to start with this slide that says, "When it's eternity here, it's still early morning on the west coast."

(Laughter.)

DR. BERG: I appreciate being able to do this by teleconference as opposed to traveling.

I wanted to show you where I come from. This is to illustrate. In the distance, you can see the outline of the United States. These are the five states for which there's only one medical school, the University of Washington. I spend a good deal of my time traveling around the region, and a good deal of that time is spent in the offices of physicians who are actually trying to deal with the complexity of medical decisionmaking. And I can tell you that the issues of genetic testing are a very big issue in our region, trying to figure out how to make sense of an increasingly large and confusing body of literature.

I'm going to spend just a few minutes giving an overview of clinical guidelines. My background is actually not in genetics or genetic testing, but in the development of clinical practice guidelines. I worked with the Preventive Services Task Force for a number of years, with the Institute of Medicine on some guidelines related to vaccinations, and I'm now also chairing a CDC panel about genetic testing that I'll get to in a moment, and an IOM panel on post-traumatic stress disorder. So my background really is more in clinical guidelines than it is in genetic testing.

Guidelines have always been with us. They really are simply preformed recommendations issued for the purpose of influencing a decision about a health intervention. And we've always had them, as long as medicine has been practiced. Professors pontificate. Textbooks give us advice. Journal articles, editorials, consensus panels, and so forth.

The problem is that in the past many of the guidelines have just been wrong. They've been well-intentioned and well-advised by an expert but they've proven just incorrect in practice.

There's also been extremely wide variation in practice which not only leads to wide variation in outcome, but wide variation in costs.

Medical literature is increasingly complex, which makes it difficult for an individual clinician to get their arms around a given clinical topic and make sense of it.

Patients are interested in more participation in medical decisions, and guidelines when published and explicit are usually publicly available which allow patients access.

There is always legal pressure to help define standards in medicine.

And finally, part of the renewed attention in guidelines is simply because we've got better methods to generate them than we've had in the past.

From a clinician's point of view, no one can keep up. The volume of medical literature is enormous and growing. Guidelines help make sense out of what can literally be thousands of articles about a given clinical topic. They help clinicians deal with complex decisions, we hope improve the quality of decisionmaking, and increasingly provide justifications to patients, payors, and even the legal system about why decisions are made the way they are.

So guidelines are potentially useful to transmit medical knowledge, to assist patient and physician decisions. They're a way to help set clinical norms. They're increasingly used in quality improvement projects in hospitals and group practices. They're used for privileging and credentialing and also can be used for payment, cost control, and medicolegal evaluation.

In the past, most guidelines were constructed using what I'd call global subjective judgment. It's a technique where you basically lock clinicians in a room and tell them to figure it out. And you really don't know much about the process that went on. Nowadays, of course, guidelines are increasingly explicit and evidence-based, and there are several hallmarks of an evidence-based guideline. It should be explicit, that is, clearly laid out. It should be transparent so anyone going back to look at it can figure out how you reached the conclusions that you did, and it should be publicly accountable. So it should be published and available not only to clinicians but to all comers.

Here are some of the characteristics that the Institute of Medicine believes should be specified when developing a clinical guideline: first of all, to be extremely clear about the clinical condition; the health practice or intervention that is proposed; the target population; the health care setting, whether a specialist setting or a primary care setting; the type of clinician, nurse, physician, nurse practitioner, physician assistant; the purpose of the guideline, whether to improve clinical care or have some other purpose; and finally and very importantly, the source of the guideline and sponsorship, that is, who's paying to have the guideline constructed.

The Agency for Healthcare Research and Quality has also specified a number of process characteristics. These are things to look for when you're looking at a clinical practice guideline. How was the panel selected? In particular, what were the screens for potential conflict of interest. How was the problem specified? Very explicitly, how was the literature strategy devised, how was the analysis conducted, how was the evidence summarized? And linking the evidence to the recommendation needs to be as explicit as possible. This is often still one of the black boxes in clinical guidelines, but as much as possible to be explicit about how you get from the evidence to the recommendation. To be clear about the clinical outcomes, and finally, the process should be sensitive to cost and practicality.

The AHRQ further described desirable attributes of a guideline. There's a separate slide on validity, but the guideline needs to be valid. It needs to be reliable so that it acts the same in each circumstance where it's applied. It needs to be practically applicable. It needs to be flexible, clear, multidisciplinary. It should be peer reviewed before publication, and it needs to be well documented.

Then finally, on issues of the clinical guidelines, here are some characteristics of validity that AHRQ recommendations. A valid guideline should be clear on projected health outcomes, on costs, on any parts that relate to policy rationale. It should be evidence-based and rigorously based on the literature review, evaluation, and on the strength of evidence.

So that's an extremely quick overview of the guidelines business. I see around the table a number of individuals who are quite expert in this who, I'm sure, could answer any questions better than I.

But I'd like to move on and discuss one particular guideline project. Again, I see a couple people in the room who are very familiar with this, and that's the EGAPP project. This is from the Centers for Disease Control and Prevention, EGAPP standing for the Evaluation of Genomic Applications in Practice and Prevention.

This is a slide that I call "Parents of EGAPP." These are some of the principal reasons I think that EGAPP was formed.

First of all, an obvious, growing availability and promotion of genetic tests. You have only to go to the Internet and put in "genetic tests," and see the many thousands of hits that you get, many of which are available to consumers without going through any sort of clinical advice.

A second parent is that clinicians need authoritative advice. This gets back to my experience with the five states for which we're the medical school. The clinicians out there in practice really would like to know whether these tests are ready for clinical use.

Finally, one of the parents, I think, is the natural evolution of these evidence-based processes that were used previously. One example is the United States Preventive Services Task Force.

Now, here are some of the challenges I see in using the standard evidence-based methods for genetic tests. First of all, as opposed to some of the conditions that the Preventive Services Task Force worked with, for example, with breast cancer or colorectal cancer, many of the conditions in genetic testing are uncommon or exceedingly rare. In many circumstances, the interventions and clinical outcomes are not well defined. The technology is evolving quickly so that the interventions -- sometimes the test characteristics change quickly, and we haven't had time to really examine the clinical outcomes in detail.

Many of the tests have inadequate sensitivity and specificity in unselected populations. They may be very effective tests in highly selected populations, but when applied in a general population, lose important test characteristics and thus have poor predictive value.

Many of the tests that we see are proposed and marketed based on descriptive evidence and pathophysiological reasoning with really no clinical trials yet.

And there is an important overlay of advocacy from various sources, but especially from industry and from patient special interest groups.

EGAPP has the CDC as its principal sponsor. It's a nonregulatory panel; that is, we don't have any inside track on any regulatory authority. We're all independent, non-federal employees and very multidisciplinary. The panel went through an extraordinary review of its own conflict of interest to make sure that we had no one on the panel who had made up their mind about genetic testing or who had financial interests at stake in any of our decisions. And we're trying very hard to make the panel evidence-based, transparent, and publicly accountable. We do have public sessions.

Our goal is to establish and evaluate a systematic and sustainable mechanism for pre- and postmarket assessment of genomic applications in the United States.

And we've spent a good deal of our first couple of years working on the methodology. Here are some of the characteristics of the methodology. Devising a method for choosing the topics was in itself a major task, given the many hundreds of tests out there. Devising a methodology for constructing analytic frameworks for our literature search strategies and for our assessment of the evidence. This particular domain of genetic testing adds analytic validity to the other common kinds of validity that you look for in testing, which has presented its own particular challenges to the panel. And finally, a methodology to properly specify clinical outcomes. Many of the clinical outcomes in genetic testing are different from the clinical outcomes that one would look for in other domains of medicine.

We've actually gone fairly far, I think, advancing the field of clinical outcomes. We're pretty far in a manuscript for potential publication that outlines four general categories, one category being the diagnostic thinking or health information impact.

Another, therapeutic choice, impact on patient outcomes, and finally, impact on the family and society.

Our work plan is to select topics, define the outcomes, and conduct reviews, and then to test the methods. The first several topics that we're examining are CYP450, HNPCC, and screening for ovarian cancer.

We also are experimenting with brief reviews where the data are quite limited. We may not be able to cover all the components in a full clinical practice guideline so the scope is narrower and not in as much depth. And our first review is a UGT1A1.

We're midway into year three of a three-year project that's extended to four, and rumor has it that it may end up being five. We hope for three to five major reviews, two to three brief reviews, publications about our methods, and finally, rigorous evaluation.

I thought I would just conclude by walking you through one topic which is fairly far advanced with EGAPP to give you a sense of how the panel works. This is the clinical scenario that we specified for our review of CYP450 testing.

The question: does testing for CYP450 polymorphisms in adults entering SSRI -- that's selective serotonin reuptake inhibitors -- treatment for nonpsychotic depression lead to improvement in outcomes, or are testing results useful in medical, personal, or public health decisionmaking? So this is the question that we hope to provide useful advice to clinicians and patients and others.

The methods that we've used were, first, to develop an analytic framework. Out of that framework, we extracted a series of key questions; around each of those key questions, then

conducted an explicit search using a standard abstract, full text, and two reviewers; assessing the quality of evidence; and putting together evidence tables, when there was enough to put into a table, which wasn't often.

And these were the key questions. First of all, the overarching question: does testing improve outcomes? A second, derivative question: what are the characteristics of the tests? What are the correlations of the tests with metabolism, efficacy, and adverse effects? Are there any known effects on management, clinical outcomes, or decisionmaking? And are there harms associated with testing?

We're not ready to release the recommendation yet, but here are some preliminary observations from our discussions so far. We found some data on sensitivity and specificity, but no studies -- and I would underline "no studies" -- directly linking testing to clinical outcomes. The studies that we have are small, poor quality, mostly cohort studies. We found no studies that directly compared alternative testing strategies, and many of the studies fail to account for all of the relevant genotypes, making it extremely difficult to combine the studies into a single clinical recommendation.

So you can tell from that that this has been quite a challenge. The panel is meeting in about three weeks to finalize our clinical recommendation out of this data source.

Here are the other topics that are currently in review: tests for ovarian cancer, HNPCC for patients with colorectal cancer. I mentioned the brief review of UGT1A1 for patients treated with irinotecan for colorectal cancer, and we've just begun gene expression profiling in breast cancer, genomic profiling for cardiovascular disease, and CYP450 profiling for pain management.

So I'll conclude with two slides. This is kind of my summary of the apparent gaps in evidence, given my experience so far in this domain. There's a gap in evidence about the prevalence of some of these abnormalities in the general population, a gap in evidence regarding the penetrance of the abnormalities into something that's clinically recognizable. There's an absence of clinical trials that compare testing and intervention strategies, an absence of studies that fully assess all the relevant outcomes, very little attention to harms, mostly just attention to potential benefits, and very little literature regarding cost and feasibility of these technologies.

Then my concluding slide, which are my personal observations. This is a large and growing question for clinicians here in the United States, both for clinicians and for consumers on how to make sense of this. We tend to be in a national attitude where more is always better and that technology is always good. I just recently returned from a meeting of the German Genetics Society in Bonn, Germany, and it's just always fascinating being elsewhere and finding such a completely different view of technology and how it fits into the social good.

We have an environment here which is relatively hostile toward regulation.

There is potential using these technologies for both benefits and harms and, unfortunately, finding limited evidence. And I'd have to say that I knew that this was going to be a problem getting into it, that our evidence base was going to be limited, but I didn't realize how limited it would be. We've chosen some very important topics for our initial few reviews which are supposed to be about as good as it gets, and yet, we're finding major gaps in the evidence in all the reviews that we've undertaken so far.

So I apologize for zipping through this, but I know that the discussion is always more interesting than a presentation. So I'd be happy for comments or questions. Thank you.

DR. FERREIRA-GONZALEZ: Thank you, Dr. Berg. This is very timely with some of the tasks that we're going to be undertaking.

I would like to ask the other two speakers to come up front, please, Dr. Vance and Dr. Richards. If you can sit at the front of the table.

I will open the committee for questions. We have a limited amount of time. So if we can have a specific question from the committee. James?

DR. ROLLINS: Dr. Berg, this is a question for you. On one of your slides, you talked about the gene expression profile for breast cancer. Currently a number of insurance companies are paying for a mamoprint, as well as oncotype dx for gene profiling for the occurrence of breast cancer. It seems like maybe the cart was put before the horse because a number of insurers are currently paying for it, but yet, this is one of the proposed studies that you're going to be looking at in the future in terms of the EGAPP goal?

DR. BERG: Yes, and thank you for that question. One of the reasons that this area particularly interested me is that I spent 10 years with the Preventive Services Task Force where the horse was long out of the barn before we got to any of the topics. There are many things that are out there that are being routinely promoted and used for which the evidence base is actually quite thin, and I was hoping that for this domain, we might have a crack at getting to some of these things early enough to help clinicians and consumers make the decision before it becomes the standard of practice without evidence.

So we're not very far into that particular assessment, but I'm hoping that we can move quickly enough to actually get ahead of it and help clinicians and consumers decide whether it's a good idea, and if so, in what circumstances, and if it's not a good idea, what are the circumstances that we should be wary of.

So I think it's an excellent question that relates to this domain in particular. There are many things out there that are being marketed that would be nice to have an assessment done before the horse is out of the barn.

DR. FERREIRA-GONZALEZ: Reed?