CLINICAL UTILITY AND COMPARATIVE EFFECTIVENESS

Clinical Effectiveness, Clinical Utility, Comparative Effectiveness: An Evolving Landscape

Gurvaneet Randhawa, M.D., M.P.H.

[PowerPoint presentation.]

DR. RANDHAWA: My charge from our chair is to go over clinical effectiveness, clinical utility, and comparative effectiveness and where things are moving. It is a fairly large set of issues and I won't be able to go into them in any depth, but I hope it will provide you with a flavor, highlight some things, and hopefully set things up for Dr. Sox to take on from there.

So, effectiveness. Many good things come from Yogi Berra. I don't know if he said this or not, but I did find it on the Web. This is the challenge with effectiveness.

The other thing that we had touched upon briefly yesterday was what is translational research. There are many steps involved in moving from a brilliant idea that has been shown to work at the bench to actually using it in clinical practice. In my perspective, there are three major areas: moving from the preclinical science to clinical efficacy, moving from efficacy to effectiveness, and then probably the hardest one, moving from effectiveness to implementing programs and using it in practice.

So, what is the difference between efficacy and effectiveness. Simply, it is the fact that whenever we perform tests or offer therapies in the average clinical practice, you don't see the same benefits and harms that you would be expecting from efficacy studies. The big question is why. As you can imagine, it is not just one factor why. There are certainly many patient factors that can influence effectiveness, and the foremost is biology. I know some folks equate genetic variation with biology, which I think is a part of it but perhaps only a major part for most things.

So, the person's age. If the studies have been done in middle-aged persons with the same results and the same benefits, will they be seen in older adults, will they be seen in children. The sex of the person.

The comorbidities. If you have liver cirrhosis, your liver is not functioning, or if you have kidney failure, how does that change the effectiveness of practice compared to studies that were done in generally healthy people. The severity of the disease has an impact, and of course genetic variations.

Apart from the biology, there are many other patient factors: adherence to the drugs or other therapies, the costs from the patient's perspective, the preferences to what therapy he or she would want, and of course, although this is not really the patient's preference, but drug-drug interactions that do occur that are not intended or studied in the efficacy trials.

I will highlight natural history, which one can argue is part of the biology, but this is a very important issue in terms of do we actually know the natural history of the disease. This is often where some of the recommendations or some of the controversies arise. How well do we know that carcinomas will progress to local cancer or progress to metastatic cancer and cause death. Some of the controversies about prostate cancer screening are a good example of that.

There is also the related issue of surrogate versus health outcomes. What is really being studied as an outcome in the efficacy trials. More often than not, it is surrogate outcomes. When we are studying surrogate outcomes, we have to have a very good indication that there is a good link between the surrogate outcome to the health outcome.

I can give you some examples from the U.S. Preventive Services Taskforce where lowering blood pressure in patients with high blood pressure, or lowering cholesterol in patients with high cholesterol, were surrogate outcomes that the taskforce felt comfortable will predict health outcomes. Lowering hepatitis C virus titers was not enough evidence for the taskforce to say this will lead to reduced cirrhosis and improved health outcomes.

Apart from the patient perspective, there are issues around the provider, the skills and training of the provider, their experience. This is particularly true for implanting devices during surgical procedures. How many have you done, what kinds of patients have you done them in.

Of course, there are provider preferences, too: what kinds of devices will you be implanting, how much time does a provider have to deliver an intervention, what is the coverage and reimbursement.

Then there are issues about the hospital or maybe the health system in general: what kind of a hospital it is, how many patients has it seen, what kinds of facilities are available. I will give you an example of Warfarin to highlight some of these issues.

In Warfarin, we know that it is an effective drug. It reduces thromboembolic events in patients who have a risk for thromboembolism. It could be somebody who has had deep inner thrombosis. It could be someone who has had or is having an issue of fibrillation and has a heart valve transplant and they are at high risk for a thromboembolic event. It is one of the most commonly prescribed. From the data I have seen, it is among the top 10 medications in the U.S.

It also has a very narrow therapeutic index. In this case, the effectiveness of the drug is measured by looking at INR, International Normalized Ratio, which tells you the amount of anticoagulation in a person at that point. If the INR level is too high, there is a risk of bleeding events that can lead to stroke and lead to GI bleeding. If it is too low, you are not really reducing the thromboembolic events in the future.

The challenges are, how well do we monitor a patient's INR, often there are drug-drug interactions or diet-drug interactions that can modify the effectiveness, and adherence.

There have certainly been trials in pharmacogenetics, but I will give you another example of personalized medicine, which is can the patient do their own INR monitoring. There have been studies that show that if you do weekly monitoring of the INR, about 85 percent of the patients will be in their target INR range, which is usually around 1.5 to 3.0, depending upon the condition. If you do only monthly monitoring, it is more around the 50 percent range.

The obvious question is, can the patient monitor their own INR at home. There was a meta-analysis done in 2006 that looked at 14 randomized control trials. Two of them were in the U.S., one in Canada, and the rest were in Europe. They had a variety of designs, all the way from those who just monitored their INR at home and then communicated those results to the provider, to those who also had a dosing algorithm to adjust your own dose based on what your INR results are.

Here is also an interesting example of surrogate outcomes and health outcomes. What was found in these studies is, for the people who were self-monitoring their INR, there is an increase in the proportion of people who have INR in the target range.

Now, the studies are reporting this differently, so there was no one pooled estimate after, but all 11 of those studies had trends in the same direction. Six of them had statistically significant results. These were small studies. Some had as few as 50 patients. Most were in the 100- to 200-patient range, which I think is an important point because the recent coag trial had patients in the same range and did not show statistically significant results for surrogate outcome.

More importantly, this meta-analysis showed that there is a decrease in thromboembolic events in these patients, a decrease in major hemorrhage, and a decrease in mortality, and fairly impressive decreases.

AHRQ had commissioned a report three years ago that came up with criteria that could be used when a systematic reviewer is looking at the published studies to see if a study qualifies as an effectiveness trial or an efficacy trial. The first one is patient population. Is the patient population in the primary care clinic setting -- that would be an effectiveness study -- as opposed to a tertiary hospital with a referral population.

The second is the stringency of the eligibility criteria, the inclusion and exclusion criteria. Most of the efficacy trials have fairly stringent criteria which make it difficult to generalize the results to the average population.

Health outcomes. Again because of the time span of the efficacy trials and often because of sample size, most of them do not have data on health outcomes. They usually focus on the surrogate outcomes, whereas effectiveness trials would be focusing on the health outcomes.

The other aspect is the length of the study. Again, it takes time to analyze for long-term events, and the effectiveness trials are designed to do that.

Another criteria is, did the trial actually assess all the adverse events systematically. Another one is sample size. Was there enough of a patient population to actually identify those outcomes. Finally, analysis.

There was a different slide set that I had created. I think this is the older one. That is okay; I will ad lib.

I don't need to go into this in detail. What I wanted to do was move on from effectiveness to utility. There is some confusion in the field when we say clinical utility. What I wanted to get across here was that there is a term called health utility used often in the health services field that looks at a patient's preference for a health state. One way of measuring it is if you are in perfect health your utility is one, given by the patient. If you are dead, obviously it would be zero, and there are numbers in between. There are different ways of assessing utility.

What I wanted to get at was that the utility itself is an outcome measure. It can be used to compare different interventions or it can be used to derive quality-adjusted life-years and disability-adjusted life-years, which are then used for cost effectiveness studies to compare the outcomes of different therapies or different treatment choices.

Where I think there is a bit of a confusion in the field is when we talk about clinical utility, where it doesn't seem to be an outcome, it seems to be more of a decision. I was looking at the EGAPP wording. Of course, a plug for Genetics and Medicine; the January issue had several papers from EGAPP. One of the papers was on methods. EGAPP was looking at effectiveness and net benefit in their definition of clinical utility, although the working groups had also considered efficacy sometimes.

The examples of clinical utility that were listed by EGAPP in the table included health outcomes, information useful for clinical decision-making, and improved adherence.

Like I said, the clinical utility is not the same concept as the health utility. It is more of a decision as opposed to an outcome measure to compare different interventions.

One point that I had wanted to make in the other slide set was that there are different factors involved in decision-making. The evidence, whether we get it from efficacy trials or effectiveness trials, and the benefits and harms are only one part of it. Another part is the added value of incremental benefits. So, if there is something new, does it provide new benefits and harms compared to something old.

Then, depending upon the decision-making context, cost effectiveness could be part of the discussion, if you are thinking about population-level decisions, individual decisions at the point of care, patient preferences, provider preferences, convenience costs, the whole shared decision-making process.

These are several other issues that come into play. It is not just simply one-on-one looking at the outcome and therefore a decision is made.

I have discussed effectiveness, so I will move on to comparative effectiveness. The issue in comparative effectiveness is, what is a comparator. What are we comparing. One is a fairly long list of clinical interventions. It could be different tests. When I say tests, it is not just lab tests or imaging tests. It could be screening protocols. It could be checklists. I'm using the term fairly broadly here. There are devices, drugs, dietary supplements, biologics, surgical procedures, counseling, and behavioral interventions, and you can go on.

So there are many different types of clinical interventions. Sometimes we are comparing one versus the other or within the same class of interventions which ones actually work better.

Some folks are defining comparative effectiveness to include health care programs and delivery systems, so one can make it broader. The only challenge is, the more broad you make the definition and the study design, the harder it is to tease out what factors are actually leading to improved outcomes.

The other part about comparative effectiveness is, what are the methods, how do we get at the information. There will be some issues about the study design. I'm sure you will hear about that later from one of the speakers. We have a fairly robust tool kit, if you can say that, for studying outcomes. We certainly need to do some tweaking. So, for doing randomized control trials, having more head-to-head trials looking at effectiveness would be needed. We already have established that this is a superior methodology.

Observational studies, modeling, systematic reviews, meta-analyses, and of course we need some work on analytic techniques that minimize bias and confounding, which reduce internal validity of the results.

One point that I wanted to get across is, there is some confusion that any evidence-based medicine principle, or I prefer the term evidence-based decision-making, equals a randomized control trial and one is not below the other. That isn't quite correct. The Preventive Services Taskforce and certainly the EGAPP Working Group have the principles of looking at the magnitude of net benefit -- so, how much do the benefits outweigh the harms -- and the certainty of that. How well do we actually know that that will occur in practice.

You can get that data from observational studies, too, but it is uncommon. The Preventive Services Taskforce has made recommendations on cervical cancer screening and phenylketonuria screening, and there are no randomized control trials on these.

There was recently an EPC report -- EPC is an AHRQ program, Evidence-Based Practice Center -- which looked at different treatments for obesity. They based their conclusions that surgery is very effective for morbidly obese people, people with a BMI greater than 40, on a very well done observational study in Sweden. Surgical methods led to reductions of weight in excess of 44 pounds, which is far superior to any medical intervention, and there was no randomized control trial data.

I think the point is, the magnitude of benefit was so much that it is very difficult to explain that by confounding and bias. Those kinds of things are not seen too often in our experience.

I will briefly go over what AHRQ has been doing in this area. There is comparative effectiveness research at AHRQ. We have had a program center since 2005, because Congress had authorized in Section 10.30 of the MMA Act that AHRQ should do comparative effectiveness research. The goal of this program is to provide the patients, the clinicians, and the policymakers with reliable, evidence-based health care information.

The Effective Health Care Program looks the effectiveness and efficiency of health care for the Medicare, Medicaid, and SCHIP programs, with the focus on what is known now and building on the previous experience of the gaps in the evidence and where AHRQ can fill those gaps. The focus is on clinical effectiveness.

The conceptual framework of how the program is organized is, there is stakeholder input in all different phases of the conceptual framework. The first step is doing horizon scanning, trying to figure out what the evidence needs are that need to be met and filled. Once we get that, there is a website for people to put in research questions. We talk to our stakeholders and get that information.

Then the decision is made at AHRQ on what is the next step. Is there enough evidence to merit doing an evidence synthesis or a systematic review, or do we need to fund a study to create the evidence or do evidence generation. Once that research is done, the next step is disseminating and translating that into practice. There are also research training and career development as part of our programmatic activities.

So, what are some of the outputs of the program. A couple of years ago, we released a study that compared effectiveness of different treatments to prevent fractures in people who have low bone density or osteoporosis. There is another example of an executive summary on comparative effectiveness and safety of oral diabetes medications.

These are executive summaries of what our EPC program creates, which we call CERs, Comparative Effectiveness Reviews. These tend to be fairly technical. Then we go to the next step of trying to create some clinically useful products. There is a clinician guide and a consumer guide that tries to make this information available in a concise, actionable form where both the certainty as well as the uncertainty of the findings are communicated.

I won't go there because I think Dr. Sox will follow up on this. There was another point that I had in the other slide set. Where we stand right now with genomics is, it is fairly easy and relatively inexpensive to get genetic information. The volume of information that you are going to get will be enormous. What we know is there is very little data on either the outcomes or the added value of these tests to our ongoing interventions. We have already heard in the previous sessions about how, with increasing life span, an aging population, increasing obesity, more comorbidities, and new technologies, health care is becoming more expensive. Genetics is likely to exacerbate all of this.

I have mentioned before that we have the EPC reports. I mentioned some of the projects on producing new outcomes in clinical decision support tools. There are some things that we are doing, but we need to do a whole lot more. I will end there.

DR. TEUTSCH: Great. Thank you, Gurvaneet. That is good.

[Applause.]

DR. TEUTSCH: You are going to be here for the day, right?

DR. RANDHAWA: Yes.

DR. TEUTSCH: You know we are running late, but I think there will be some questions. If you are here, they will come up as we go along. So, thank you, and thanks for your adaptability with having the wrong slide set available to you.

I think it is apparent to everybody that the reason there is so much attention at the federal level to this is, this is one of the few things that are likely to provide some solutions to the rising health care costs. So, the work is getting cranked up.

One of the people who has played an enormous role in this for many years and certainly is again at this time, is Dr. Harold Sox. He has been chairing the Institute of Medicine's Committee on Comparative Effectiveness Research Prioritization. That group was tasked with recommending the particular comparative effectiveness studies the government should undertake with the ARRA funds.

Harold earned his medical degree from Harvard and has served on the faculty at Stanford and Dartmouth. He has most recently been the editor of the Annals of Internal Medicine. I understand, Harold, that we are getting to the last month of that tenure.

DR. SOX: Four more weeks.

DR. TEUTSCH: But who's counting. I'm sure that there are some important next steps which I don't know about, but Harold has made some important improvements in the Annals of Internal Medicine to bring this kind of information to clinicians to help them practice better.

We were hoping, Hal, that you would be able to talk to us about the comparative effectiveness agenda from the IOM perspective on where this field is going and give us some hints about how genomics might fit into all of this.

I will remind the committee that we did send a letter to Hal on behalf of the committee. Again, it mostly emphasized the importance of including genomics on the comparative effectiveness agenda.

It is always wonderful to see you here, Hal. We appreciate all your leadership over many years in bringing good information to clinicians so they can make better decisions.

Future Directions and the Role of Genomics in Comparative Effectiveness

Harold Sox, M.D., M.A.C.P.

[PowerPoint presentation.]

DR. SOX: Thank you, Steve. I want to say first that everything I'm going to say today is in the public domain. The reason for emphasizing that is that Institute of Medicine reports are embargoed until they are released. I don't want anybody to interpret anything I say as reflecting the content of the report, so everything is in the public domain. I will try to be as careful as possible on that score.

CER, Comparative Effectiveness Research, and the promise of this is really thrilling to doctors. It is a focus on making better decisions. I can't think of a program of research that has more of a focus on something that is so important to patients and physicians, as well as researchers who work in this field.

Steve has already said something about the ARRA and the role of CER in it. The only thing I would add is that the funding timeline is that the money has to be obligated by the end of next calendar year, although I gather it can be spent for considerably longer than that. We are not limited to really short-term studies. On the other hand, we would like to have some short-term studies get done, get published, and make a difference so as to build public support for this type of research.

Now, definitions are really important. They tell you what is and could be funded with CER funds. Our committee spent a fair amount of time trying to conflate the other definitions that are out there into something that is short and sweet and covers everything.

Our definition is two sentences: "The generation and synthesis," meaning both original research as well as summarizing the research that is out there already, "of evidence that compares the effectiveness of alternative methods to prevent, diagnose, treat, monitor, and improve delivery of care for a clinical condition." You can see it is a very broad field of topics to be included under this umbrella. "The purpose of CER is to help patients, clinicians, purchasers, and policymakers make better-informed health decisions."

Let's briefly talk about what is unique about CER. It is unique, I believe, because it includes all five characteristics that are listed here. I have circled the first three because I think they are really the most important for us to keep in our heads. The first is direct head-to-head comparisons of alternatives, treatments, tests, or whatever, any of which might be the standard of care.

Second, the study population should be representative of clinical practice.

Third, the research should be patient-centered in that it should help physicians and patients to tailor the choice between alternatives to the specific characteristics of that patient, using data gathered by the physician and offered by the patient. It has a broad range of topics, as we have already noted, which includes the delivery of health care, the translation of research into practice, and a broad range of potential beneficiaries.

I want to say an extra word about the patient-centered concept. Let's suppose we have a randomized trial that shows that Treatment A is better than Treatment B. Sixty percent of patients respond to A but only 50 percent to Treatment B. Nonetheless, since 50 percent of the patients responded to Treatment B, it is clear that it is by no means an inert substance.

If all you knew about the patient was that they were like the patients in this trial, then you should prefer Treatment A.

Is it possible that some patients actually should have chosen B despite the fact that most patients got better on A. Can we identify those patients in advance and steer them in the direction of the treatment that they are most likely to respond to. That is an intriguing research question that I believe should be an important one in the research agenda. That is just a personal view.

Now I'm going to try to give an example of the principles of comparative effectiveness research to genetic testing for diabetes susceptibility. I made these slides pretty late last night and, in a fit of madness, didn't include the reference, which was to an article in Annals of Internal Medicine, the journal that I edit, in its April 21st issue, for those of you who want to follow up on this.

Let's see how things go here. Steve Goodman is going to come along to pick up the mess that I leave in terms of the analytic side, so I know I'm safe in venturing out on a limb.

Here is the background. Genome-wide association studies have identified a number of loci associated with type 2 diabetes and a number of SNPs associated with each of those loci. The purpose of this study was to examine the joint effects of genetic loci and conventional diabetes risk factors. In other words, to compare conventional risk factors' ability to predict who is going to get diabetes with a combination of genetic information plus conventional risk factors. So, what does the genetic information add at the margin. That is clearly a CER question.

The study, which was done by a group mostly based at the Brigham and Women's Hospital in the Harvard School of Public Health, attempted to predict the onset of diabetes in women, taken from the Nurse's Health Study cohort, and men, taken from the Health Professional Follow-Up Study. It was a subset of these patients who agreed to give blood for testing.

It was a case-control study in which the cases were those who developed diabetes and match controls who did not develop diabetes over a period of about 20 years, during which time the participants were contacted by the study every couple of years to see if they were reporting the onset of diabetes. The exposure in this case control study would be these genetic loci and the SNPs and conventional risk factors.

The goal here, then, is to calculate the odds ratio for exposure. In other words, the frequency of these SNPs in cases versus controls. By a wonderful mathematical trip, this is mathematically equivalent to the odds ratio for being a case that is having diabetes or developing it given exposure versus no exposure. Any of you can prove that to yourself with mathematical manipulations that you learned as a freshman in high school.

The conventional risk factors they examined included BMI, physical activity, and energy intake, because they did dietary assessments in these participants periodically. They calculated a genetic risk score, GRS, where, basically, the more SNPs you had, the higher your risk score. They had both the strictly additive model as well as one that weighted different SNPs differently. The goal then was to have a multivariate model to predict diabetes risk.

Here are the main results. They divided the participants into quintiles of equal size according to their genetic risk score. The numbers in blue represent the odds ratio for developing diabetes. None of these patients had diabetes at the outset. You can see that there is a nice dose response curve. The higher the genetic risk score -- in other words, the more SNPs that were associated with the development of diabetes -- the higher the odds ratio for developing diabetes.

This was, importantly, adjusted for a number of risk factors for diabetes. It implies that the presence of these SNPs make an independent contribution to predicting diabetes incidence over and above the conventional risk factors.

So far so good, but now we go on to another way to look at this, which is the ability of this information to discriminate between people who will develop diabetes and those that won't. To do that, you calculate an area under the ROC curve. That is not shown in the next slide.

Believe it or not, I couldn't retrieve the figure from my home computer because I didn't have the sign-in to retrieve it. It is crazy. Four weeks to go. I may still do it.

The ROC curve actually gives you the probability that a person who is destined to develop diabetes will have a higher score than somebody who is not destined to develop diabetes. As it turned out, the area under the curve for conventional risk factors was 0.78, which means the probability that somebody who is destined to develop diabetes will have a higher score is almost 80 percent.

If you add in the genetic risk score, it is 0.79. Basically, it doesn't make any contribution, or at least any clinically important contribution, to discriminating between people who will develop diabetes and those who won't, which would be important for targeting programs to try to reduce the incidence of diabetes through the use of behavioral change as well as Metformin.

So, why does the genetic information add so little discriminatory power. One possibility is that in the statistical analysis there is colinearity, which basically means that the genetic factors influence the diabetes risk through the conventional risk factors and so, in effect, don't really add any information.

Another possibility is that the prediction is so good with just the conventional risk factors that genetic information can't add much.

Still a third possibility, which may be the best one of all, is that the area under the curve is really a poor measure of discrimination. Some of you who are hip on this stuff will know that there has been a big flurry of interest in what are called reclassification indices, which basically measure the ability of a prediction rule or prognostic rule to move somebody from a medium risk either to a high risk or to a low risk. These may turn out to be better measures of the addition of extra information like diagnostic tests in predicting the future, which will really be a very important development, I think, for CER. We are going to see a lot more of these reclassification indices.

Let me say a few words about our committee. As Steve in his introduction said, the ARRA mandated a study by the Institute of Medicine that had to report by June 30th, which was exactly 19 weeks after the President signed the bill into law. It was to include recommendations on national priorities for CER. In other words, conditions or research questions to be addressed with the CER money that you heard about earlier. In addition, they mandated that we consider input from stakeholders.

We built on the experience at AHRQ in our approach to trying to get stakeholder input. First, we held an open meeting at the National Academy of Sciences building, where we heard from 56 presenters in seven hours and had a really good opportunity to ask questions of them. It was really a highly satisfactory meeting which held its audience, both people who weren't on the committee as well as people who were, really quite well. As these types of meetings go, they are always very rewarding. You come away with a really good, warm feeling.

In addition, following AHRQ's lead, we did a Web-based survey that was open to anybody. Mostly it was health professionals and organizations of health professionals that made recommendations. We asked them to give us their top three condition-intervention pairs in order of priority. We had over 1,000 unique respondents and over 2,000 nominations, of which a number were duplicates entered by somebody who really wasn't in the spirit of things.

Here are some of our priority-setting criteria which were outlined on the website. This is the information that we really asked nominators to identify as one of the reasons for making their nomination.

In addition, we paid a lot of attention to trying to get a balanced portfolio of topics so that we didn't leave any important area completely high and dry. For that we developed several criteria for trying to balance our portfolio and paid a lot of attention to that during our discussions.

The next steps are that the report now actually is in the review process of the National Research Council of the National Academies. We hope that we will be able to deliver our report on time in a couple of weeks.

I'm now going to turn to a question that a lot of people are wondering, which is, in health reform legislation, will CER be there. If so, what form is it likely to take. To do that, I turn to the important white paper issued by the Senate Finance Committee several weeks ago, A Call to Action: Health Reform 2009. The language here is basically the language of the report.

It first says that a number of respected panels had called upon Congress to create a national entity charged with conducting CER-type research, including one from the Institute of Medicine, in which I participated.

They go on to say the plan would create a new institute charged with identifying the most pressing gaps in clinical knowledge. From that language you can imagine something new is going to happen.

The proposed institute would be private, nonprofit, with a board of governors representing both the public and private sectors. It would be created as an independent entity to remove the potential for political influence on the development of national research priorities. Now, whether this will come to pass is anybody's guess. This is what the Senate Finance Committee was thinking about. In an address on Tuesday at the Brookings Institution, Senator Baucus reaffirmed his preference for this arrangement.

The institute should not only recommend areas of inquiry, it should produce research. It should be able to contract with federal agencies that have bureaucracies set up to issue requests for proposals and evaluate them and generate reports based on them. It must also have the flexibility to deal directly with private researchers as well as through government agencies.

Very importantly, the institute should be open to public interest and transparent in order to maintain the integrity of the research, just as this body is open to the public and functioning entirely out in the open.

Most importantly, the institute should be subject to rigorous oversight of its finances in order to maintain the public trust. These new endeavors would need an adequate and stable source of funding. Since the research would benefit all Americans, it seemed reasonable to the Senate Finance Committee to levy a small assessment on private health insurers as a way of ensuring a steady flow of dollars that would not be subject to the annual appropriations process. That is what the Senate Finance Committee has in mind.

Finally, just a word about public attitudes towards CER. Scott Gottlieb, who is a deputy commissioner of the FDA, wrote a very negative op ed in The Wall Street Journal representing one point of view that emphasized the potential harm of doing better research.

[Laughter.]

DR. SOX: He was echoed by Rush Limbaugh.

On the other hand, the American public, as you will see in a second, seems to like the idea. I'm now going to refer to a national poll commissioned about two months ago by the Herndon Alliance. This is the part to read. This is the statement that the respondents were supposed to react to. You can see basically that a total of 73 percent favored or favored very strongly this statement and only 17 percent were against it, with 10 percent not being able to decide.

Interestingly, they framed the question two different ways and assigned them randomly to respondents. In one version of it, it had costs in it. In the other, it didn't have costs. Maybe this just reflects the fact that people didn't read it very carefully, but the strength of preference was the same whether or not cost was included in the framing question.

I will end by restating the promise of CER, information to help doctors and patients make better decisions.

[Applause.]

Question-and-Answer Session

DR. TEUTSCH: Why don't we take one or two questions for Hal. This is terrific. Hal, I hope you can stay because we hope to have more discussion later. Jim, then Sam.

DR. EVANS: I just have a quick question. What arguments do people make against this? I'm trying to think of some but can't.

DR. SOX: I can't, either.

DR. EVANS: I will call in to Rush Limbaugh.

DR. SOX: Yes, that is right. Sam.

DR. NUSSBAUM: Hal, again, congratulations on supporting all of this research, leading the IOM effort. As you mentioned on Tuesday, Peter Orszag also believes that comparative effectiveness research done right will really play a key role in bending the curve on cost.

The question I have is -- and it sounds like this is embargoed and you can't mention it -- of the 1,000 people who responded on the survey and the 2,000 ideas, did genetics rise high in the domain of what people want to look at, or was it more likely, based on the public hearings, focused on common costly illnesses like cardiovascular disease?

DR. SOX: You are right, Sam. I really can't answer that, or shouldn't answer that and won't.

DR. NUSSBAUM: Just another point. The elephant in the room, of course, is cost. People have used the issue of cost and not looking at cost in creating concern, both on the very politically right and on the political left, actually. People have been concerned that this would fly in the face of personalized medicine and it would lead to in fact rationing of care for unique populations.

You are as knowledgeable as anyone in this space. Do you think that is a concern? Not whether you think the public thinks, but do you think that it would actually cause that harm?

DR. SOX: Speaking personally, the short answer is we clearly need to know about the value that we get for the resources that we are expending on patient care. I worked for the American College of Physicians, which issued a position paper which we published that came out very strongly for including cost effectiveness information basically as part of the CER effort. We had an editorial by Gail Valinsky [ph] and Alan Garber [ph] commenting on that issue. Both basically agreed, by the way.

As everybody knows, the words "cost" and "cost effectiveness" are really toxic in this town. We will just have to see what happens.

MS. WALCOFF: I just have a quick question on if you are considering liability issues. I thought it was really important, the notes you emphasized, on using comparative effectiveness research in addition to the physician's discussion with the patient and what is best for that individual patient, the real patient focus.

Suppose a study shows that Product B is generally better for most people but the physician thinks that Product A would be better for this individual patient. Is there a concern that, depending on what that physician is basing that decision on, that might expose him or her to some kind of liability if the research is more limited on the benefits for that particular subgroup or that particular patient? Is that factored into the comparative effectiveness research protocols?

DR. SOX: I'm actually embargoed from saying anything about the process that we went through and our discussions, so I really can't say whether that issue came up or not during the discussion.

Speaking just for myself, I think that we need to understand a lot more about the degree to which malpractice concern actually plays a role in doctors' decisions to, for example, get diagnostic tests under circumstances where the probability of their changing care of the patient are very low. It is surprising how little research you see on that subject. We don't see very much of that at our journal. I wish we did.

DR. TEUTSCH: Julio, and then we will need to take a break.

DR. LICINIO: I had a question. You brought up the very important issue of the autonomy of this entity and the idea that it should not be part of the NIH or a public entity because of the fear of political influence. If you put it in the private sector, essentially make it independent but with a private component, and fund it apparently exclusively by the insurance companies, would that create another type of potential influence?

DR. SOX: What are you thinking of?

DR. LICINIO: In terms of setting agendas, for example. If something is of interest for an insurance company, can they lobby and put direct or indirect pressure for what should be a topic of study?

DR. SOX: What leverage would they have? The money that is funding the enterprise is coming from a tax that exists because it is a law.

DR. LICINIO: There may be people on the board that have alliances to them.

DR. SOX: The Senate Finance Committee, as I remember, said something about how there should be both private and public sector representation on the governing board. Presumably, there would be open declaration of people's financial relationships. Because the meetings would be occurring, and I'm hypothesizing now, just like this one, out in the open with anybody to comment and to see if people are ruthlessly pushing their particular financial advantage, it would be unlikely that that would lead to the group as a whole making a decision reflecting one person's lobbying effort.

DR. TEUTSCH: Part of it was federal.

DR. NUSSBAUM: Actually, the health plans, about two years ago, suggested this type of funding, a tithe, to lead to sustainable financing. A lot of this is being worked out in additional legislation being proposed in the House and Senate, but it is one of many funding sources.

I think the theme that Hal is pointing out is the public-private partnership theme to this because everyone benefits, as opposed to, just historically, a government agency looking at these issues, where the focus might be actually more on CMS beneficiaries or others.

DR. TEUTSCH: Thanks so much, Hal. This was a terrific presentation. Thanks for all your work over many years. All the best as you move on to the next phase.

Please, if you are staying, we are going to have a panel at the end. We will have the chance to revisit this with all the speakers who can stay with us.

You should have received the draft of the memo to David Blumenthal. If you have any comments, would you please get them to Sarah before noon? If you think it needs discussion, get back to her. Otherwise we will see to finalizing it. Thanks.

We will take a 10-minute break and reconvene before 10-to. Thanks.

[Break.]

DR. TEUTSCH: As we continue our discussion on clinical utility and comparative effectiveness, our next speaker is Dr. Michael Lauer from NHLBI. He is director of the Division of Prevention and Population Science. He is a cardiologist by training and completed his work in cardiovascular epidemiology at the Framingham Heart Study. He joined the staff at the Cleveland Clinic in '93. During his 14 years there, he established a world-renowned clinical laboratory research program focused on diagnostic testing and comparative effectiveness.

We have asked Mike to talk from the perspective of NIH because, as you have heard, NIH is playing an increasing role in the comparative effectiveness world. Here again, he can't speak to the specific priorities, particularly as they relate to the ARRA monies, but he will be talking about the focus on the role of genomics research and comparative effectiveness from the NIH perspective.

Welcome, Michael. It is always good to see you.

We look forward to what you have to say.

Role of Genomics in Comparative Effectiveness Research: NIH Perspective

Michael Lauer, M.D.

[PowerPoint presentation.]

DR. LAUER: Steve, thank you so much for the invitation. I'm going to briefly review a number of areas of interest to the NIH in comparative effectiveness research. First, I will review the history of comparative effectiveness research at NIH, a little bit about the many definitions of CER, the impact of the Stimulus bill on CER, how NIH activities on CER are organized, and then a few closing thoughts about the opportunities and challenges that the Stimulus bill present to us.

The first question is, do we really need to have CER. I think, as you have heard from the speakers before, it is quite clear that there is a need for evidence.

This is an interesting study that was done by Sid Smith, Rob Kaliff [ph], and colleagues, where they went through all the guidelines and recommendations that have been released by the American Heart Association and the American College of Cardiology over the last 25 years.

They made a number of interesting discoveries.

The first is that the number of recommendations being given to doctors has increased dramatically. You would think that is great, but the number of recommendations that are actually based on solid evidence, that proportion has actually gone down. Most of the new recommendations that have come out have been based on soft or absence of evidence.

The second thing that they did was they looked at those recommendations that are currently active and classified them as being based on Level A evidence, Level B evidence, or Level C evidence. Level A evidence means real evidence. It means multiple randomized trials. Level C evidence means opinions or consensus or "expert" opinions.

What was found was that only 11 percent of currently active recommendations in cardiovascular medicine are based on Level A evidence, whereas 50 percent are based on Level C evidence. Fifty percent of the recommendations and current guidelines are based on expert opinion only.

Now, the NIH has a longstanding history of comparative effectiveness research. We have been doing this for decades. In fact, in this week's New England

Journal of Medicine, the lead article is the main results of the BARI 2D trial. This was a major comparative effectiveness study that compared revascularization versus medical therapy in over 2,400 patients with diabetes. It also compared insulin sensitizing therapy versus insulin provision therapy in these patients with diabetes.

It found, actually, that there were no differences. The outcomes were just as good with medical therapy as with revascularization and just as good with one kind of diabetes therapy as with another. This is just an example this week of a major comparative effectiveness study funded by NIH that came out.

The study that Hal Sox mentioned earlier this morning that was published in the April 21, issue of Annals of Internal Medicine was also funded by the NIH.

Here are some other examples of major landmark comparative effectiveness studies. We have drug versus drug. The upper left-hand corner is the CATIE trial that compared different drugs for schizophrenia. The middle one is the ALLHAT trial that compared different hypertensive drugs in people with hypertension.

The upper right-hand corner is screening versus usual care. This was a big trial which I will show you in a moment. It compared the use of a screening test for preventing deaths from cancer.

The bottom left-hand corner is lifestyle versus drug. This is a diabetes prevention project that compared lifestyle versus drugs and found that lifestyle actually did a better job of preventing the onset of diabetes.

In the lower right-hand corner is an example of a drug versus device trial. This was a trial comparing Amiodarone to defibrillators for prevention of sudden cardiac death in patients with heart failure. It looked like the defibrillators did better. These are just a small set of examples of many comparative effectiveness studies that the NIH has funded over many decades.

Here are two examples of trials that have just come out this year. This is screening versus usual care for prevention of deaths from prostate cancer. This was a trial that involved 77,000 men. They were randomized to get a screening PSA and digital rectal exam versus conservative management. What was found was that patients who were randomized to the screening arm had more cases of prostate cancer diagnosed. That is good. That is exactly what you would hope to find.

However, there was absolutely no difference in the rate of deaths. In fact, actually, the death rate from prostate cancer may have been a little bit higher in those people who were randomized to screening. This is a huge comparative effectiveness study funded by NIH.

Here is another one, a smaller study that compared two different types of surgery for patients with coronary artery disease and left ventricular dysfunction. One type of surgery involves bypass. That has been done for a long time. The other kind of surgery involves removing a portion of the ventricular wall and then putting the rest of the heart back together. This is an operation that has actually been fairly popular for about 10 to 15 years and was gaining in popularity.

This trial compared these two approaches. It turns out that there was absolutely no difference in the outcomes. Probably a simple bypass operation alone will do.

Here is an example of a trial that we are doing right now that directly hits upon genetics. This is called the Clarification of Optimal Anticoagulation to Genetics trials, or the COAG trial. One of the major reasons I went into cardiology was that I loved the acronyms. Cardiovascular trialists are very good at this.

[Laughter.]

DR. LAUER: This trial is going to compare two strategies for dosing Warfarin. It is a randomized trial looking at patients who have an indication for being on Warfarin for at least three months. They will be randomized to one strategy in which genetic test results will be used to determine dosing, and the other strategy will be based on the clinical algorithm only.

There are two genes here. One is called the 2C9 gene, which affects the disposition of Warfarin. The other is the VKORC gene, the Vitamin K Epoxide Reductase gene, and that affects the target of Warfarin. It turns out that these two genes are fairly common and have strong associations with the Warfarin response.

We have a very large infrastructure for doing comparative effectiveness research. Again, one that has been around and has been developed for many decades includes clinical trial networks, cooperative groups, disease registries, and the HMO Clinical Research Network.

This is a network that is being funded through the National Cancer Institute and the NHLBI in which data are being extracted from electronic medical records of over 10 million patients.

There is a consensus development program for evidence syntheses. The National Library of Medicine has a Center on Health Services Research. CTSAs, or the Clinical Translational Science Awards, are relatively new over the last few years. The idea of this is to bring community collaborations into clinical research.

There is now active collaboration between NIH and FDA on post-market surveillance. Within the National Cancer Institute, there is integration of the SEER cancer surveillance data set with CMS. There are huge infrastructures in place, with lots of people with lots of expertise in areas of comparative effectiveness research.

Now, with this new interest and the new legislation, we have had to struggle with many definitions. Hal briefly alluded to those. There are lots of definitions. Here are a couple of them.

The CBO definition, the Congressional Budget Office definition, came from Peter Orszag's report in

December of '07, in which he said that CR is a rigorous evaluation of the impact of different options that are available for treating a given medical condition for a particular set of patients. Such a study may compare similar treatments, such as competing drugs, or very different approaches. I'm just showing you some examples of studies funded by NIH that would fit that.

The FCC is the Federal Coordinating Council. This is the council that was put together by the new Stimulus bill to oversee the federal government's efforts in comparative effectiveness research. The first time I saw in an Email we are going to have to consult the FCC, I thought, what does the FCC have to do with this? I felt too dumb to ask.

[Laughter.]

DR. LAUER: Anyway, the FCC is using this definition, or at least it was using this definition when I made this slide. "Conduct and synthesis of systematic research comparing different interventions and strategies to prevent, diagnose, treat, and monitor health conditions."

I think there are some interesting points here.

One is that there is conduction of research and there is synthesis of research. Also, this goes beyond treatment. It also involves prevention, diagnosis, and monitoring. It also points out that the purpose of this kind of research is to inform patients, providers, and decision-makers about which interventions are most effective for which patients under specific circumstances. Mike McGinnis at the IOM has a great line for this. It is "the right treatment for the right patient under the right circumstances in the right setting."

Here are some common themes that exist across these definitions. One is that there is some kind of valid comparison. We are comparing something against something else. The second is that the research is focusing on effectiveness as opposed to efficacy. Effectiveness means that we are dealing with the real world. These are real patients being seen in real circumstances in real practices. We are dealing with available options. In other words, not drugs or devices that are only available under IDEs or that are highly novel or virtually nobody is using it.

There is also a focus on real outcomes. One way of thinking about real outcomes is, real outcomes are those that real patients and real policymakers really care about. Real outcomes would include length of life, quality of life, prevention of major clinical events like heart attack, stroke, hospitalization, diagnosis of cancer, and cost.

The Stimulus bill has presented the government with a unique opportunity to focus renewed attention on comparative effectiveness research to the tune of $1.1 billion. NIH is getting $400 million. AHRQ is getting $300 million. The Secretary is getting $400 million. Much of the impetus for this bill comes from the Congressional Budget Office report that Peter Orszag put together.

Peter Orszag, of course, as you know, is now the director of the Office of Management and Budget. One thing that he loves to focus on is the plot there on the right showing variations in health care spending across the United States. I don't know how many of you read Atul Gawande's fabulous essay in the current issue of The New Yorker in which he pointed out that McAllen, Texas, which I will admit I had never heard of before, now has the distinction of being the most medically expensive town in

America.

The point is that there are huge variations in resource use in medical care across the country. Yet, these variations in resource use do not appear to be related to outcome. Elliott Fisher published a terrific paper in Annals of Internal Medicine in 2003 in which he looked at that. There has been a variety of analyses after this that all show the same thing.

The NIH, in response to the Stimulus bill, has formed a coordinating committee. This is chaired by Betsy Nabel, who is my supervisor and the director of the National Heart, Lung, and Blood Institute, and Dr. Richard Hodes, who is the director of the National Institute of Aging. I'm on that committee.

We have been charged with a number of responsibilities, including determining how we should best use the Stimulus funds, how we should best collaborate with sister agencies and in particular with AHRQ, how we should put together our portfolio analyses of just exactly how much CER we are doing and of what type, how we can best communicate and disseminate our CER findings, accelerating research through existing mechanisms and new programs, which I will talk about in a just a second, and then considering the agency's long-term charge for CER.

Again, NIH has been doing comparative effectiveness research for a very long time, for many decades. We see this as an opportunity to jump-start a new pace of CER, but something that should go way beyond the two-year span of the Stimulus bill.

We plan to obligate the $400 million in ARRA support for a variety of activities. One is peer-reviewed meritorious grants. What this means is that over the past couple of years there have been a number of investigator-initiated grants that came in that got good scores but, because of our budget limitations, we were unable to fund. We are now going to be able to fund these grants. In fact, yesterday I was in a meeting of the coordinating committee and we went through a number of the grants that we are considering funding.

The second is supplements to current research. These are people who already have grants or contracts, providing them with some additional money. This is actually a relatively small part of the NIH spending plan.

Challenge and grant opportunity grants. How many people in this room either sent in a challenge grant or know somebody who sent in a challenge grant?

[Laughter.]

DR. LAUER: How many people in this room missed meetings because of that?

The challenge grants are two-year, $1 million opportunities in a variety of areas. One specific area was CER. We received 21,000 challenge grants, of which 1,700 were specifically in CER. We are now in the process of reviewing them, and it is going to be a busy summer.

The second big area are the grant opportunity grants. The grant opportunity grants are two-year grants for more than $1 million. We did issue RFAs specifically in comparative effectiveness research. I don't really know how many we have received. I know in NHLBI we have received about 50, but that is an incomplete count.

Now, the next area are contracts. Many of our trials are funded by contracts. We will be exploring within the next two years areas in which we can enhance those trials. Funds will be awarded based on peer review, scientific opportunity, and potential biomedical and public health impact.

Now, there are a number of challenges that the Stimulus bill has presented. Scientists, even highly-driven scientists, are not used to two-year timetables, so this rapid timetable has presented some interesting challenges for us and for the scientific community.

One of the major worries that the scientific community has is what we are referring to as the cliff. That is the cliff that is going to happen in two years when this bolus of spending suddenly disappears.

Two-year funding mechanisms are unusual. Most NIH grants are four or five years. Many of our contracts are seven to eight years. There is a political context within which all this is happening. You have heard some of it this morning. The term "cost effectiveness research" gets a number of people very uptight.

There is the question of economic impact. Now, there is economic impact of the Stimulus funding, which is that we hope that by providing this money to researchers, universities, small businesses to a lesser extent, that we will be either creating jobs or retaining jobs.

There is also the question about the economic impact of comparative effectiveness research. There are some people who feel that this is going to be the answer to all of our health care woes and will dramatically cut cost. There have been other estimates that have suggested that the impact will be much more modest.

Interagency contexts. This provides a great opportunity for agencies to cooperate more with each other than they have been. We have had some great examples of interagency cooperation. There are a number of research projects that are jointly funded by NIH and AHRQ. We funded a major comparative effectiveness study on emphysema surgery in which CMS issued a ruling that they would only cover the operation as part of the trial. That is another great example of cooperation and collaboration between agencies.

What will be the long-term effects of a one-time bolus infusion. We don't know. The level of accountability is at unprecedentedly high levels. We keep getting reminded about this constantly. We have been told, for example, that we are not allowed to have communications with registered lobbyists unless that occurs in writing. There is real worry that registered lobbyists will be trying to directly interact with NIH staff on specific projects or applicants, and we have been told that we have to be very careful.

Pressure on review functions. The NIH normally gets about 77,000 grant proposals per year. It is estimated that this year we will get 115,000. All those people who are writing grants are also being told that we expect them to review, and we are hoping, of course, that we will be able to do this review in both an expedited but also fair and objective way.

Stay tuned. The comparative effectiveness research train is moving very fast. I want to thank you again for the opportunity to be here.

DR. TEUTSCH: Great. Thank you.

[Applause.]

Question-and-Answer Session

DR. TEUTSCH: I know Steve is going to put up his computer. Why don't we see if we have a couple questions. I know this is a timid group. Sheila.

MS. WALCOFF: On the registered lobbyist limitation, I just saw I think it was about a week and a half ago that the White House Counsel's Office had expanded that to lobbyists and non-lobbyists. I don't know if that is more restrictive or less restrictive on you, but I just wanted to alert you to that in case it hasn't gotten down throughout the departments.

DR. TEUTSCH: Other questions or comments?

[No response.]

DR. TEUTSCH: To what extent do you foresee the genomics portion playing a role in all of this?

DR. LAUER: I think it is going to be fairly huge. As you know, much of the genomics work right now has been primarily in the area of epidemiology. We have put genomics data from Framingham and we are about to put genomics data from WHI and MAISA [ph] and other big trials into publicly available databases. This has been used primarily for studies of mechanisms of disease and epidemiology of disease.

NHGRI has an initiative to incorporate genomics with clinical trials. We have all these clinical trials. We have funded many clinical trials. We have biological specimens from tens of thousands if not hundreds of thousands of people. DNA can be extracted. We can now do genotyping for much lower prices than we used to. It is actually now realistic to talk about genotyping 10,000 or

20,000 people who are in a trial.

I can't talk about specific proposals, but there are a lot of them. I actually saw yesterday the list of projects that we are considering funding, and there are some real good ones. The clinical trials area, I think, is another big area.

The other is that investigators are getting interested in doing genomics-based trials. COAG is one example, but we have seen proposals from investigators where they want to actually test an interaction to see whether or not a treatment is more likely to work in a group with a certain genotype as compared to a wild-type genotype. They are actually proposing trial designs and giving us these trial designs to look at.

My guess is that, particularly as the cost of genotyping is going down, we are going to be seeing more and more of these trials and we will be funding them.

DR. TEUTSCH: Great. Thanks so much, Michael. I know you have to get off to Cleveland, so thank you for visiting with us before you have to take off.

Our next speaker is going to focus on some of the challenges going forward with methodologic issues and doing these kinds of studies for a very fast-moving field. Until relatively recently we have had a fairly constrained set of processes for doing this, and you have heard some of them today about systematic reviews, trials, and somewhat in the observational study range. These present some real challenges for a field that is changing as fast as this one.

We asked Dr. Steven Goodman, who has been heavily involved in thinking about these issues for a long time, to talk to us about where this field might go. We are deeply appreciative that he could come today. As you heard from Hal Sox, Steve also serves on the Annals as the guru of all things methodologic, as well as the doer of all of these things.

It is a real pleasure to have you here to help us think about where this is going and how we might think about all of this. Steve, welcome.

Future Directions and Developments in Research Methodologies

Steven Goodman, M.D., M.H.S., Ph.D.

[PowerPoint presentation.]

DR. GOODMAN: Thank you very much. I never was introduced as a guru of anything, so I don't know if I can quite live up to that.

I do have to divulge a conflict of interest here. I have worked with Gurvaneet on a project recently, and he knows that I have completely eschewed the use of the terms "clinical utility" and "clinical validity" as hopelessly confusing and unclear. I don't know if that banishes me from the room, but I'm not a big fan of those.

I will focus on some aspects of this. Predicting is always hard. I think that is another Yogi Berra quote. Prediction is hard, especially when it is in the future, something like that.

DR. TEUTSCH: Niels Bohr.

DR. GOODMAN: Oh. Thank you very much. Yogi and Niels were very close friends.

[Laughter.]

DR. GOODMAN: I'm going to be focusing on a very, very small piece of that, not specifically on CER but in the genomics realm. I did want to follow on Hal's promise that I would come after him and help out some of the technical points.

This is the miracle of having computers with your whole life on it and all your talks. I thought I would just show this slide, which shows the relationship between population classification and individual classification. What you see here are two populations that correspond to a biomarker.

This is the distribution of the biomarker and the probability of the number of people who have the biomarker of some arbitrary value. This corresponds to two populations, non-diseased and diseased, where the odds ratio related to that biomarker was 1.5. Here the odds ratio is 3.0. That is actually pretty large for most risk factors in most epidemiologic domains.

You see that no matter where you cut these populations your sensitivity and specificity is going to be awfully bad. These populations are almost right on top of each other. The reason that we get this discrepancy between what we think are large effects and what are extremely poor effects has to do with the focus on individual classification.

What we are usually interested in, until now, in the epidemiologic realm is distinguishing between populations. We can increase the sample sizes and we can make the means of these two populations arbitrarily precise, and we can see that little difference. That doesn't mean on an individual level that we can discriminate very, very well.

In order to have biomarkers or genes or predictions that have anything close to the sensitivity and specificity we need, we have to have the equivalent of odds ratio of 25, 70, which you never see. That explains that phenomenon that you saw occurring of genetics often having very little predictive power when it looks like they have some contribution to the prediction equations. That is why that is happening.

This is just an ROC curve. This is an ROC curve of a factor that has an odds ratio of 2.0. You can see it is very, very poor, with the diagonal having no information.

That is just a little background. That was just for Hal. I couldn't resist.

Here we go. These are things that have been identified as cancer risks: electric razors; broken arms, but only in women; fluorescent lights; allergies; breeding reindeer; being a waiter; owning a pet bird; being short;

being tall. If you have escaped all those possible classifications, there is hot dogs and having a refrigerator. We are all at risk.

Now, this isn't genomics specifically, but I could show the same sort of thing 10 times over in the genomics realm except you wouldn't laugh. You would say, oh, that looks interesting. The names would be KET45, 47Z95, and things like that.

It is a big problem. We are generating these reams and reams of relationships and we don't know what they mean. Here are the problems and the conundrums. You already know this. This is what I will be talking about some of the approaches to.

Often, a little background or mechanistic information helps sort out the noise from the signal in the discovery of genomic associations of putative clinical importance. In addition, the pace of discovery is much faster than the pace of evaluation. I should have put "discovery" there in quotes. The finding of statistical associations is not really a discovery, but too often we treat it as such.

Then these things are put on the table for evaluation. When we are looking at evaluation measured in human lifetimes, that obviously has to be slow. We have to be very, very careful about how we allocate our human experimental resources. Obviously, it generates a large number of potential genetic, genomic, metabolomic, and proteomic combinations.

I didn't want to make you wait for the solutions. I have all the solutions here. We will go through them. Of course, these are not absolutely solutions but they are the beginnings of approaches. There are many more than I am going to list on the slide, but this is just going to be a few things that I talk about.

[No.] 1 is new clinical trial models. I'm going to focus on Bayesian adaptive designs that allow for rapid introduction and prioritization of new therapeutic genetic combinations. I'm going to talk very briefly about two trials that are ongoing, the I-SPY2 and the BATTLE trials, which are actually examples of this.

We need to reexamine regulatory standards and guidance that impede novel evaluation approaches such as these. I have also been told that FDA has a requirement that when you are doing a cancer trial that one of the agents actually be an established cancer therapy. That makes it very, very difficult when you are developing targeted therapies that individually might have no effect but work synergistically, knocking out two steps in the same pathway. That is very, very difficult to get approved as a single agent.

We need support for development of tissue repositories that link clinical data and long-term follow-up from RCTs. This is a huge lost opportunity and often the only way we can get rapid results. This, of course, was the way that instruments like Oncotype DX was validated on NSABP clinical trial data from the '80s.

Actually, there are very, very few resources like that. Every clinical trial that ends without long-term storage of the specimens and clinical follow-up, which is the key, is a potential waste of that original investment. We actually have the power to be able to test many of the things that we are developing if we would start investing in this. In many trials that aren't of the NSABP type that information gets lost. We might have the tissues, but we don't have the long-term clinical follow-up. We don't have enough of it.

We need to improve methods to identify biologically and clinically relevant signals with high throughput results. I'm also going to put in one of my soapbox items, improve methods and establish standards for reproducible research. I will just talk very briefly about that.

Let's talk about the Bayesian adaptive designs. Bayesian adaptive designs are trials that change based on prospective rules. These are not anything-goes trials. They are very rigorously design. They changed based on prospective rules and accruing information, focused experimentation, and the most promising or informative directions.

Almost everything about these trials can change as they go on. You can change the sample size, the randomization scheme, and the accrual rate. You can drop or reenter arms or dose groups. You can explore combination therapies or doses. You can stop early for success or terminate early for futility. Most importantly, you can adapt to responding subpopulations. You can actually change endpoints from clinical endpoints at the beginning of the trial to surrogate endpoints at the end of the trial if you see during the trial that they are correlated.

All the rules that many of us have been taught about prespecification and rigidity of design, these are actually artifacts of a traditional statistical method -- you don't want to get me going on that -- that doesn't allow for natural and common sense learning. Bayesian designs allow us to do this. The methodology is all there. We need to do a lot to get it into practice, but it is being championed by folks from MD Anderson, particularly Don Berry, who has taken the lead in getting this into practice.

Here are two trials that are currently in the planning or execution phase. I would say that at MD Anderson they have literally done hundreds and hundreds of these.

This is I-SPY2. It is an adaptive breast cancer trial design for neoadjuvant chemotherapy. That is chemotherapy in women with large localized tumors before surgery. This is to shrink the tumor to allow for a higher chance of a definitive cure.

The problems that are trying to be addressed by this design are that clinical trials take many years for both the development and evaluation of new therapies and often ignore tumor heterogeneity, and also that the use of biomarkers for both prediction of patients who will respond to drugs and for the early assessment of that response are badly needed for more informed, faster, and smaller phase three trials. You will see that they do an amazing amount in the one package of this trial.

The basic design of this trial is, women who are HER2-positive are randomized to Paclitaxel plus Herceptin, plus or minus a new drug. Then they go on to traditional chemotherapy. Actually, there is a missing arrow here. Women who are not HER2-positive, basically the same thing, except they don't have Herceptin, obviously. They go on to traditional anthracycline and cyclophosphamide. They have MRIs and tissue samples early on, and then they have definitive surgery.

This does not actually do justice to what the trial is all about. That is more on the next slide.

It has two goals. One is to evaluate new therapies in patient subsets on the basis of the biomarkers. The second is to test, validate, and qualify new biomarkers as drugs are tested. I will talk about how they classify those biomarkers.

Regimens that show a high Bayesian predictive probability of being more effective than standard therapy graduate from the trial with their corresponding biomarker signature. If a particular therapy and a particular biomarker subgroup looks like it is very highly promising, that actually leaves the trial for testing in the phase three setting. Regimens are dropped if they show a low probability of improved efficacy. New drugs can enter as those that have undergone testing are graduated or dropped.

This is a learning trial system. We talked about the learning health care system. This is the learning clinical trial system, like we would all think common sense would dictate.

The setting, as I said, is neoadjuvant. The eligibility I have already mentioned. The endpoint is pathologic complete response.

There are three biomarker classes. There are the standard ones like HER2, estrogen receptors that are used for patient eligibility and randomization. Then there what they call the qualifying biomarkers that have great promise but are not yet approved. They are used for the subgroup analysis. Then there is the exploratory biomarkers, for which there is very preliminary data. These come and go within the trial.

This is a list of the eligibility criteria for drugs. They start with a certain panel of drugs, but new drugs can come in, as I said, as those drugs come out. It is what you would expect. It has to be compatible with standard therapy. It has to have some reason to believe it would have some efficacy. It has to target any of the key pathways that are associated with the biomarkers. The drug must be available.

This is what is called the BATTLE trial, short for Biomarker Integrated Approaches of Targeted Therapy for Lung Cancer Elimination. Cancer easily competes with cardiology. This is a design paper that just appeared in Clinical Trials last year. This, again, is a trial where we have multiple biomarker groups. The biomarkers here are EGFR, K-RAS, VEGF, and Cyclin D. Basically, if you are positive EGFR, you are in Biomarker Group No. 1 regardless of the others. It actually proceeds downward like that until you are negative on all. This is what they predict the population will look like.

All five groups are then randomized to these four therapies. So there are 20 possible groups here at the start, with a minimum of 20 per group that is going to be tested. The randomization probabilities change as the therapy-biomarker combinations are more or less successful. It is just like the other one. They can graduate, they can stop, and the arms are dropped and more combinations added depending on what the results are.

In Bayesian adaptive designs, experimentation is a continuous process. More patients are treated with better therapies. Trials can be shorter, but not always. External or patient-specific information can be incorporated.

When are Bayesian designs more efficient. We see that they are more flexible. They are more efficient when the result is consistent with prior evidence and the evidence is permitted. That is no small thing. We are not used to actually formally incorporating prior evidence into the interpretation and design of the trials, again because of a statistical paradigm that is now 80 years old. How many other technologies do we use that are unchanged in 80 years. We should be embarrassed.

Bayesian adaptive designs are also more efficient when design adaptations minimize unneeded experimentation -- that is, by dropping subgroups or arms -- when there can be a smooth transition from one phase of research to another, and when surrogate endpoints are informative and occur before the definitive ones.

When are they not more efficient. When the result is inconsistent with the prior evidence or that evidence isn't permitted, then the boat has to sit in each tub on its own bottom. Then you can't really borrow evidence. That is the only way to get more information from what looks like less. Somehow you are gathering and synthesizing evidence from multiple sources. If those multiple sources are seen to be not relevant or in conflict, you don't get any more efficiency. You just have to learn from the evidence in front of you.

When there are no subgroups or arms that can be curtailed, when you can't seamlessly go from one phase to another, and when surrogate endpoints are in fact not informative, then you are stuck with waiting until the end.

I will tell you that adaptive designs are no small thing to implement. They require intensive up-front planning and simulation of the designs. This next point is, these trials are really important in exemplifying a very sophisticated data infrastructure that allows accrual and integration of almost all clinical, genetic, proteomic, treatment, imaging, and outcome information in near real time.

If you don't have this, then you can't make decisions that change the trial. You can't just wait two years and then break the code and do the analysis. This is happening in real time. We are accountable for high-quality data management on a time scale that we are not always used to in clinical trials.

What is holding us back? Flexible, user-friendly software for the statistics, design, and data management. It has to basically be built anew for each trial. Few statisticians and clinical investigators have experience in designing and carrying out these trials. It does require a lot more up-front planning time, and people like getting their ideas into the protocols and in to the IRB and getting started in weeks or a month, and you can't do that with these. You get the payback on the back end, not on the front end.

Also, an unfamiliarity of government regulators with Bayesian designs holds us back. This is changing but still very real. I don't really blame them. The academic community itself is not that familiar with them.

Again, some of the solutions. I have talked about new clinical trial models, support for development of repositories. I won't read these all again. I will talk briefly about the reproducible research model so you at least know what that is. This was written about in an article by some of my colleagues in the American Journal of Epidemiology. I have to show Roger Peng's picture here because this is really his life's work, and it is not mine. I have to say more than just his name, so that is Roger, who works on this.

A reproducible research model is something, again, that we haven't really seen and may not be used to. In a sense, the data, the methods, the documentation, and the distribution are all part of one document. It is a fused document that has the data and all the code embedded, but it looks like a paper that you would read. You can actually live reproduce all the analysis. You could change one point and change all the figures and all the data. It is a new way and a new standard of how research is presented. It first came out of very, very technical proposals in the computer programming literature and is now starting to see broader and broader application.

The current data-sharing model is basically you share or you don't share. Authors put stuff on the Web or they don't, or they send it to you or they don't. It might be in a journal's supplementary materials. In genomics, we do have some central database for a variety of domains, but it doesn't really solve this problem completely. Readers have to get the data, download it, figure it out, and get the software and run it. That is no small thing.

Now, the data-sharing model actually involves issues of intellectual property that are very much like intellectual property rights for software and other things. There are ways you can constrain how the data can be used. I didn't put that slide up here, but it is much more complex than just giving people data or not. It is a mutual partnership between the person who has the data and the person who might use the data. There are all shades of gray between total use and total non-use, which is the model right now.

This is the pathway where we have our measured data down here. Then we have our analytic data set, then our computational results, and then we generate sometimes hundreds of figures, tables, and results. Then we merge these with text and we get an article at the end of the day, and that is what we see published in the Annals or wherever.

The reproducible research model allows the reader to go all the way back here, where all of these things are actually fused within the single document. It allows for a lot more transparency.

I have to show this since Hal is here. We are trying to move this into the clinical research arena. We have made some baby steps. We can't require our authors, obviously, to do anything like what I have been describing, but I bring it to your attention as a direction in which I think we are going to be moving over the next five, 10, or 20 years. What a research article is going to look like in the new electronic age I predict has to be very, very different. It can't be a PDF of something that appeared in paper.

Reproducible research can improve the transparency and accuracy of published research and enhance the value of post-publication peer review. For the people in this room what is important is it makes questionable results and methods easier to detect and correct. It accelerates and improves reanalysis and data synthesis. These are all things of interest, I think, here in the genomic realm, where there is a lot of spurious stuff being generated.

Here are the same solutions. I think I have touched on almost all of them. I don't have a set of possibilities there, but I only had 15 minutes to talk about it. That is another few days. I think I will stop there and take any questions. Thanks.

[Applause.]

Question-and-Answer Session

DR. TEUTSCH: Thanks, Steve. That is great food for thought for this. Why don't we take a couple of questions for Steve. Hopefully he will be able to stay for some of the discussion, too. Jim.

DR. EVANS: That is really fascinating. Given the multiple arms, I imagine you have to look at conditions for which you have a large number of people. I think about that because there was a study a few years ago that showed that there was essentially one randomized clinical trial in the entire field. I think part of that is not excusable and part of it is because we deal with uncommon things.

DR. GOODMAN: We don't worry about power as much because we ask a fundamentally different question. We don't ask, are these treatments statistically significantly different than each other. The question might be, what is the probability that this treatment is the best. That is a different statistical question than saying, I can statistically discern this from the bottom one or from the next one. When that probability gets high enough, it goes out.

The other thing is that the information being used for that contrast is far more complex than a simple binary contrast. If you have 20 in this group and 20 in that group, you are also sharing information from that therapy being used in all the other groups and the hierarchy within that group. So, your effective sample size is larger than the 20. This is where the Bayesian formal modeling produces effective sample sizes. This is what is called borrowing strength.

It is the same thing we do when we look at patterns. When the dose goes up, the response goes up. That is exactly what I would expect because of X and therefore I believe it because of X. If you didn't know anything about the dose, if you just labeled those dose categories as A, B, C, and D, you couldn't make that inference. You have automatically, in a sense, created information by knowing that things are ordered. Some of these are modeled a priori.

There are two ways to answer the question. The effective sample size is larger than you see in the subgroups, and the statistical questions you ask are somewhat different and require less information to answer definitively. You also have a coherent way of expressing degrees of certainty. You may choose to act in the phase two setting in graduating to a phase three setting when you are 85 percent sure, and you have a vocabulary to say that. There is nothing in traditional statistics that allows you to say I am 85 percent sure, no P values, none of the technology.

You might say 80 percent. When it is 80 percent sure I'm going to graduate this to a phase three trial. It is the phase three trial that then provides more definitive information. These are screening trials or filters that move you on to the next phase. I think that is the best way I can answer that.

DR. TEUTSCH: Thank you, Steve. I hope you can stay to be part of the discussion as we figure out where we are going from here.

Let's pass the baton to another one of ours, Marc Williams, who is known to all of us. In his day job he works for a terrific organization, InterMountain Healthcare. They have done an enormous amount of work in translating information on effectiveness into quality care. Hopefully, it will help us understand how we go from what was new information into actually helping people.

Impact of Comparative Effectiveness Findings on Clinical Practice

Marc Williams, M.D.

[PowerPoint presentation.]

DR. WILLIAMS: Yogi Berra did say, "I have never said half the things I have said." I would note, though, that when I was asked by Steve to do this talk that another great American came to mind, and that is Mark Twain, who said, "It is better to remain silent and be thought a fool than to open your mouth and remove all doubt."

Now, those of you who have interacted with me in this or other settings would probably be shocked to know that I was even aware of that quotation, much less ever contemplated it. However, I think it is important to say up front that I'm not sure I'm the best person to present this. The person that has really worked for 20-plus years on this at InterMountain Healthcare is Brent James. Brent has been involved nationally in the recent discussions on comparative effectiveness research.

The things that have been going on at InterMountain Healthcare have not necessarily been labeled with the rubric of comparative effectiveness research, and so I thought I would at least present what I know, having gone through Brent's advanced training program. I have shamelessly stolen some of the slides from that program without his permission.

We tend to think about this as more quality improvement or improvement. To reduce comparative effectiveness to half of a table on a slide is probably ridiculous, but I think we have heard this morning that the definitions are evolving. Hopefully it will be a little bit easier to settle on the definition of what is a genetic test.

Methodologies are diverse. I'm not going to recapitulate this, but obviously we just heard about a couple of methodologies that I haven't even represented on the slide here.

Quality improvement is really primarily management of processes. It also uses a variety of methods. It is not primarily a research tool, but I hope we will demonstrate to you that it can result in impressive improvement in care and that that improvement in care is in fact able to be disseminated.

I did want to define what a process is. It is a series of linked steps, often but not necessarily sequential, designed to cause some sort of outcomes to occur, transform inputs into outputs, generate useful information, and add value.

Of course, a lot of this comes from industry, specifically the post-World War II Toyota model and work by Demming and others that have really helped to transform in industry what quality means. We found that these concepts actually will operate in the health care arena.

To do process management, you have to start with a knowledge of what are the processes that you are dealing with, understand the processes aggregate to create systems and that these processes interact, and know that there is clearly variation in terms of the operation of the processes. It does require, much as we heard from the last speaker, a system for ongoing learning. What we want to try and do is to build a system to manage processes, and then ultimately, if that is a rational system that works, you get what results as quality improvement theory.

When we are defining and measuring outcomes in medicine, we can roughly aggregate these into three buckets. One would be characterized as physical outcomes. These would include medical outcomes such as complications, therapeutic goals, morbidity and mortality, et cetera. Some of these are patient outcomes, like functional status measures and perceptions of outcome.

I think it is important to recognize a flaw in much of the research that is published about patient outcomes. Many of the patient outcome studies that are published are actually physicians' interpretations of what the patient outcomes actually are, as opposed to the patients telling you what their outcomes are, a not so subtle but important difference.

There are also service outcomes relating to satisfaction for patients and families, referring providers, and other customers. It includes access.

Sheila had asked earlier about liability. It is interesting that medical liability operates more in the service outcome realm than it does in the medical outcome realm. If you seriously tick off a patient, you are much more likely to be sued than if you don't, irrespective of what their medical outcome is.

Now, the other thing that has been raising a lot of dander in the discussion about comparative effectiveness research is the whole issue of cost. However, cost outcomes are really an outcome of the clinical process. There are lots of costs that can be counted, but our experience has been that these are inextricably linked with physical outcomes. You cannot say, we are only going to look at medical outcomes, we are not going to look at cost outcomes. You can't take them apart. If you look at medical outcomes, you will necessarily be looking at cost outcomes, even if you don't actually report them.

What I thought I would do is to give you some examples of things that we have done. I'm going to have to really distill all of the hard work that has gone into these different projects and hopefully get across some key points about how things work and leave it at that.

Now, this was one of the first major projects that was rolled out relating to clinical care. This was an extubation protocol in the post-cardiac intensive care unit. These are patients that came in for cardiac surgery. They were transitioned into an intensive care unit. They were intubated and then they had to be extubated before they could move out to the acute ward.

As with any study, you need to know what the lay of the land is. There was a baseline data collection for approximately 18 months. What was identified here was that the mean time to extubation in hours was approximately 25, but you can see here that there is a huge confidence interval around this and huge variation in the process around this mean line.

Now, the intensivists and pulmonologists that were working on this ultimately were breaking down the process. They recognized that there were 240 independent variables that were at work that could lead to information to be presented to the physician to make a decision about ventilator management. I think most of you would agree that if you have 240 variables it is a little hard to construct a randomized control trial to control 239 of them and study how the impact of one would really do this.

The solution that was decided upon was to use a computer-based protocol where the physician was presented with information that was thought to be most relevant to the immediate decision on ventilator management. They could choose to accept that instruction or reject that instruction. All of the decisions were captured and then, on a weekly basis, all of the research groups got together and talked about what decisions were being followed, what decisions weren't being followed, and the protocols were adjusted. This was done in an iterative process over a period of time.

This was then rolled out in a trial. You can see that within literally a month after turning this on the mean time to extubation was reduced to slightly over 10 hours, with dramatic reduction in variability. Additional adjustments of the protocol were done, and then this was the final production version that was rolled out that ultimately resulted in extubation times of just under 10 hours with the range of confidence intervals essentially existing between seven and 12 hours.

Basically, this is a proof of principle that you can take extremely complex clinical processes and distill them down and result in significant patient outcomes.

Here are some other tangible outcomes that we can look at in terms of length of stay. We reduce the length of stay in the ICU, we reduce the length of stay in the acute care setting, and we reduce the total hospital length of stay.

Then this is an example of some procedures. This is arterial blood gases prior to initiation of the protocol. Each patient would experience approximately 12 draws. This was reduced to two draws after initiation of the protocol. The total cost of the hospitalization was reduced roughly by about $3,000 in 1994 dollars, which I think now would translate to approximately $7 million. I may be slightly off on that.

Here is another example. This was recognition of the evidence for patients with acute MI that did not have a contraindication that they should go home on a beta blocker. As in our baseline measurement, we were doing this successfully about 54 percent of the time. The process was broken down and a change was made. The change involved the discharge process, the discharge nurse, and the final order set. It was turned down, and we went from this 57 at the initiation in a month to 98 percent.

This also shows something typical of quality improvement which is called holding the gain. You can see how we drifted down after initiation of the protocol. This is very typical because processes and systems have inertia. We tend to return to what we were used to doing. Tweaks of the protocol had to be done at points two and three. Since that time we have been able to manage the process such that, on average, about 97 to 98 percent of the eligible patients are obtaining beta blockers at discharge. We did this to all cardiac discharge medications: beta blockers, ASARBs, statins, antiplatelet.

I wanted to show you an example of something that we commonly fall into in medicine. Here are our baseline measurements with the different values, and here are the national standards. You can see that we were performing at or above national benchmarks with the exception of our antiplatelet therapy. Now, in many situations we would say, good job, we are best in class save for statins, we are doing better than anybody else, and this is great. We compare ourselves to others.

We have taken to calling this the cream of the crap approach because we shouldn't be comparing ourselves to others that are also doing a lousy job. We should be comparing ourselves to the theoretical best practice. By ignoring the national data and essentially initiating these discharge protocols, you can see that we were at or above 90 percent on all of these measures. Again, all of these were achieved within one month of turning on the protocol.

Now, this is great, but this is clearly a surrogate outcome. We are assuming that better compliance here is going to result in that. We have actually developed systems to be able to capture this. We looked at mortality one year before and after the protocol, so pre- and post-. In congestive heart failure, our mortality dropped from 22- to 18 percent, which results, in our patient population, in 331 people being alive that weren't alive a year before. In ischemic heart disease the absolute drop in mortality was less but still resulted in 124 people alive. We had 455 total between those two.

Then you can look at similar data relating to readmissions, where we avoided nearly 1,000 readmissions in the year immediately following turning on the protocol. So these are true health outcomes, things that are meaningful to physicians, to patients, and to administrators.

I should say that one of the transformational activities that occurred in our institution is that at the hospital board meetings the treasurer's report does not come first, as it does in most health care systems. Something relating to actual patient outcomes is always presented first. We hear frequently, "No money, no mission," but the reality is if we are not paying attention to the mission, we shouldn't be getting any money.

Here are the cost outcomes. I should say that these are not trivial to obtain. Hospital accounting systems are not designed to track where we are experiencing cost savings. I can also tell you that if you are not in an integrated health care system that has health plans and hospitals and outpatient all integrated under one roof where you can get a handle on all these data, it is almost impossible to do this type of accounting. We basically had to develop a radically different way to do cost accounting to accomplish this.

Essentially, the fast-track extubation protocol resulted in savings to date of $5.5 million. We have experienced with these top 11 interventions across, as you can see, a wide variety of clinical areas. We had $20 million of improved cost structure, and we have had an additional 30 successful clinical projects. We have yet to have a clinical improvement project that has been successfully implemented that hasn't in fact saved money.

Will this work with genomics. We have heard a little bit about this trial. It is referenced in some of the reading materials in your packet. This is the CoumaGen trial that was done by our cardiovascular group at InterMountain Healthcare. It is a prospective randomized study of 200 patients. We were able to turn around the genotype in 48 minutes so we could use the information for initial dosing of Warfarin using a developed algorithm. We used a short-term follow-up of one month using surrogate outcomes.

We did find some differences in the genotyped patients. The initial dose was closer to the stable maintenance dose. This is not a big surprise because the literature is replete with examples showing that if you use this information you can better predict the final dose. We had fewer and smaller dose adjustments. There were fewer INR measurements, which did result in some cost savings. We did find that wild-type patients generally required larger doses, again not a big surprise given that the recommended starting dose is due to averaging across wild-type and patients that carry variants.

We did not find differences, however, in time in the range for the group as a whole, although pharmacogenomic guidance was better for wild-type individuals. That, at least for me conceptually, was a bit of a surprise. That is, the wild-type patients were getting better benefit from this, and those that had multiple variants, which we would expect. Of course, we were not powered to detect true differences in health outcomes of interest, although the time in the range is a reasonable surrogate measure.

We also captured in parallel -- to my knowledge, this is the first time this has been done in a prospective real-time fashion along with the prospective clinical trial -- an economic analysis where we captured all costs associated with that and were able to do cost accounting. I don't have time to present that information, but it was presented and will be published.

Why did we not find a difference. All of our patients were managed by an anticoagulation clinic. We use clinical process management in our anticoagulation clinic, so we have superior time in range compared to benchmarks. That meant we set up the field so it was going to be harder to detect differences in the first place because the patients were better managed.

It raises some interesting points to consider from the perspective of comparative effectiveness. Should a system invest in a robust anticoagulation clinic using best processes rather than genotyping. Would genotyping be more appropriate in a rural setting.

Think of a point-of-care genotype. You don't have the resources in a single two-doctor practice where they have to initiate Warfarin in some circumstances. You can't have an anticoagulation clinic there. Would it make more sense to use the genotype so that in that setting you would be more likely to get to the right result quicker and reduce results. I don't know; we will have to test that.

Could INR monitoring be optimized. Gurvaneet presented some of the data around home monitoring, which I find to be very compelling. The clinical processes applying that to standardized dose adjustments, which we have also done in our chronic anticoagulation clinic, have resulted in much superior time in range.

I think sometimes we see this being dismissed as cookbook medicine. I like to go out to eat. I like to think that my favorite chefs are actually using the same recipe, or close to it, every time I go in there, that they are not just making it up as they go along. In some ways, it is not an apt metaphor to begin with, but I would contend that the protocol-driven work that we are doing is not equal to cookbook medicine.

The process that we use involves a multidisciplinary team. We select high-priority care processes. We do evidence-based reviews to identify best practices. We then actually put the proposed guidelines out to the full range of practitioners who would be exposed to the guideline to get their comments and suggestions. We open up the guideline into a clinical work flow. We actually refer to guidelines in our place as shared baselines.

Clinicians are free to vary based on each individual patient based on their own individual judgment. The difference is we capture the outcomes from each of those decisions so that we can learn. When we refer to a learning health care system, this is one of the key components; that is, to have the systems in place where you can capture outcomes resulting from different decisions so that you can learn as you go along.

We have to measure. We learn. We eliminate professional variation, which is my preference versus your preference based on what we learned, in my case, 25 years ago and probably haven't updated since that time. Yet we retain responsiveness to patient variability, the idea that patients do vary. They vary around a number of different things, sometimes biologic, sometimes preferential, but that is okay.

The first rule is that whatever guideline we come up with, it is wrong. We put that clearly on everything. This guideline is wrong. The intent is that we are going to learn from it and we are going to get it right over time. It is a rapid learning, rapid cycle improvement. Some people refer to it as building the airplane while you are flying it.

No protocol fits every patient. More importantly, no protocol perfectly fits any patient. We would be more concerned about a physician where we looked at their practice and we found that they were absolutely following protocol 100 percent of the time. That would be a red flag to us because that implies that that physician has turned their brain off.

A concept from industry that we really think that this relates to is called mass customization. If you go to order your laptop, you can pick and choose exactly what you want to do. The manufacturing processes are very standardized, but you can rapidly customize and get a laptop that is built specifically for you using processes that are standardized with very low variability and very high reliability. The shared baseline then allows us to focus on small subsets of factors that are unique for individual patients.

These are the 10 to 15 percent of patients that really need the thought and intensity because there is something that is truly different about them. It concentrates our most important resource, which is our bright physicians and other providers, where they can really have the greatest impact on those patients.

I don't know how many of you actually manage anticoagulation clinics, but I can tell you from what I have been told that it is the bane of most internists' life. These are just miserable. It is a lot of time and there is very little reward.

Our physicians that manage our chronic anticoagulation clinic have extremely high satisfaction because they are only being asked to work on those patients where there is some really challenging clinical problem with managing their anticoagulation, which is what we all went into medicine to do. We didn't want to do a little bit of this, a little bit of that. That is all handled automatically at a much higher level than we can. Our satisfaction is actually quite high in our physicians practicing in this environment.

The protocol is really a tool that manages complexity. It retains the art of medicine because we are not forcing people into protocols. We are saying we think this is the baseline that you should start from but you need to use your best judgment to manage that patient. It actually improves productivity. We have data that demonstrate that our physicians are more productive, which they can either translate into higher income, because they see more patients, or they can translate into more family time because they can go home early.

We want to do all the right things all the time. We only want to do the right things. We want to do it every time with grace and elegance under the patient's knowledge and control.

I guess the question that I was left with after I did this is, is this really comparative effectiveness research. It is clearly comparative. I hopefully have demonstrated that we are measuring effectiveness. Where the problem comes in is with the research piece. I know, from talking with some of my colleagues that have tried to get some of this work published, that at least outside people that are looking at this are somewhat reluctant to say that this is research. Whether this would fall into some of these newer research methodologies that we need to have more exposure to I don't know.

I think the important thing, though, is that there is clearly knowledge here that should in some way, shape, or form be disseminated to improve care. I think that these approaches will work for personalized medicine. In fact, in our system we think that they will be absolutely necessary to realize benefit from personalized medicine. That is the basis of our internal strategy to promote translation and study impact.

I would recommend to you, under Tab 6, the brief commentary article by Garber and Tunis which addresses this issue much more eloquently than I. Thanks.

DR. TEUTSCH: Thank you, Marc. I think, whether or not this is comparative effectiveness research, this is a good example of how a group can take what we do know -- I think the cardiac things are a great example -- and actually then make sure that they get to patients and improve processes so that the technologies get to the right patients at the right time and improve outcomes.

A couple questions for Marc before we get everybody back up here and we get into a discussion?

[No response.]

Committee Discussion

DR. TEUTSCH: Hypoglycemia is beginning to set in. Why don't we invite all of our speakers who are still here, and hopefully many are, to join us up here at the table.

What we have now is some time to talk about where we want to go. This is one of our priority topics that we identified. It is clearly an area where a lot is going on. There is a lot of momentum. What we should be discussing is what do we want to do from here. What are our opportunities, and how can we play a constructive role in moving this field forward and getting better understanding of the value of genetic and genomic testing in clinical practice.

I will open it up to our panel and to all of you. Dr. James Evans, I knew I could count on you.

DR. EVANS: Marc addressed this, to some extent. I was wondering whether anybody wants to pitch in on where we go once we have shown with comparative effectiveness research that something is better. We are all too familiar in medicine with the old adage that doctors aren't really trainable. We know what to do in many cases and yet it isn't done very often. What do you think are the best ways of making sure those things are adopted?

DR. WILLIAMS: I will take the first shot at that because it is something that our system has really specialized in. I think that doctors aren't educable. I think they are trainable. There is a subtle but important distinction there which is probably of greater humor to those of us that grew up in the dysmorphology world.

The reality is that there are several things that have to come together to allow rapid translation into practice. One is the recognition that a problem exists. Second is the demonstration that there is a better way. The third is to understand, really, the biggest issue, which is the work flow and education pieces.

We know from physician post-graduate education that the traditional approaches to education have a very low level of effectiveness in terms of actually changing practice. Really, what you need to do is to present the relevant information to the physician immediately at the time that they have to make a decision, which is why you hear me continually harping on the idea of just-in-time, point-of-care education. I have to make a decision. I need to know what the best decision is.

A lot of the care guidelines and processes that we have running operate in our electronic health record environment under an info button. If a physician goes in to order a test, there is an info button that will present. If there is a relevant InterMountain guideline, the summary will pop up to them immediately. In real time, within seconds, they can get that piece of information that they need to hopefully make the right decision.

Also, with an electronic ordering environment, you can constrain certain decisions or request that certain additional information be presented. You can do that without suffering problems of alert fatigue relating to the idea that every time you try and do anything you are alerted to something. We have seen that in the drug-drug interaction world. That has been a spectacular failure, for the most part. So you have to recognize that.

The second piece is really understanding how physicians do their work and integrating that at the proper time. If you can match the right information at the time that the physician needs to do that decision, I think that obligates the use, for the most part, of electronic health records. It can be done by paper, but it is much more complicated to do and it is much harder to disseminate it across a large system. If you can hit those two things, then you can get very rapid compliance very easily.

The third thing to recognize is that it is not always the doctors that are the key person in the process. For that discharge medication process it was the discharge nurse that was managing the discharge order set that was the key individual. We actually removed the doctor from the process there and were able to achieve the high level of compliance with demonstrable improvements in morbidity and mortality.

DR. DALE: I have a couple of questions for Steve. I enjoyed your talk. I would be interested in your comments on how your Bayesian approach fits to analyze what is happening in Salt Lake City. Can it be analyzed in terms of group sizes, mathematics, and certainty of the answer?

The other question I have is, you mentioned some value associated with tissue banks. I would be interested in further comments about that.

DR. GOODMAN: I was on to that. I also wanted to answer Dr. Evans' question from my own perspective on it.

Obviously, the science of what makes doctors do what they do is very complicated. They always say if you want to understand the man, look at the child, or the woman. If we want to understand why doctors think the way they do, let's just look and see how they are educated, all the way back to the preclinical and undergraduate days. Virtually all the education is focused on basic biomedical processes. They have to take physics, chemistry, a whole host of sciences that none of us actually remember. They don't have to take economics or statistics.

The fact is that physicians are not equipped to be lifelong learners. I'm in an elite academic center, and I can tell you our fellows and our faculty do not understand the literature that they read. They understand the biology of it. They understand the mechanisms. They do not understand the statistics. They don't have a fine sense of the weight of the evidence provided by the designs and the results. The same sort of judgment they have developed in the clinical setting they do not have for the very literature that they are supposed to learn from.

In some way, this is a profoundly different source of authority of knowledge in medicine that is not derived from knowing how things work in the individual patient. Physicians don't have access to it. To the extent that they are educated in the preclinical years, they are taught with a whole host of cues and models. As soon as they get out of that clinical epidemiology course, it is not important anymore. They go on the rounds. Are they called to account? No. Do they have to read papers and do anything but spout what the conclusions are? Basically, no.

We see it in papers that are submitted by very high-level researchers. We see this throughout medicine. This is not a language that they are familiar with in terms of incorporating it into their practice. They have to be taught on the back end, when it is hopeless. We have spent eight years acculturating them to a different source of authority.

I know it is being done. In fact, there was just a report that came out last week about premedical requirements and such. We are constantly trying to change the medical curriculum, but if we want to have one reason why doctors do what they do, let's see what we teach them.

It is too late in the process. Actually, I don't want to say that. We do train clinical fellows in this, but it takes years. It takes years. It can't be done in a short course. That is what I wanted to say. Do you want to add something to that, Harold, before I go on? You were raising your hand.

DR. SOX: No.

DR. GOODMAN: On the second question, even though I waved the magic Bayesian wand, I don't want it to appear like magic and that we can't do many of the things that I was saying could be done in this particular context using traditional methods. By far and away, the most important things are asking the right questions, setting up the right experiment, and everything you were talking about.

That said, it is conceivably possible that there are ways of incorporating Bayesian approaches to make them either more flexible or more powerful. You have to look at the guts of any particular experiment.

It is the information-sharing issue that is key. That can be brought to bear on that process. Maybe it could be made a continuous learning process where the experiment never formally, in a sense, ended but new protocols were brought in. In the same way that we have QI with a cyclical improvement, you could have, as I was describing, a cyclical experimentation process. There are examples of this that have been done.

I would always say that looking at any design through a more powerful and common sense methodology might improve it. How much it could improve it is very, very hard to say. It could be 1 percent or a home run; I don't know. I do know the area that I highlighted is an area where there has been particularly high yield.

With respect to the tissue banks, it takes funding. I forgot to say when I listed my solutions multiple times that each of these the NIH has the power to ameliorate with more focused funding. When we have a five-year grant for a clinical trial where all funding ends for any clinical follow-up or support, maybe we should be thinking of a certain percentage that is maintained for every one and consolidated within the institution for doing long-term follow-up of many people who are enrolled in the clinical trials, where that is indicated.

You have to have, ideally, a centralized resource for the tissues. You have to have the linkage to the long-term outcomes. This is all part of a lot of the informatics work that is going on. You need support for patient contact for all these things. If the funding ends after five years, then, effectively, the information ends after five years.

This is being done in many domains right now piecemeal. I think it has to be taken on as a major national initiative to not squander the resources that we have put literally millions into building and then we let lie fallow.

DR. WILLIAMS: There is a really important point that Steve made there, and that is the idea of the continuous learning. It doesn't necessarily compartmentalize itself well into what we traditionally define as a research project. I think that is really critically important.

There is another protocol that we have developed on glucose management in the intensive care unit that we not only have gotten up and running in all of our different intensive care units but have also built on either a Web-based server or laptops. We have disseminated that to multidisciplinary investigators across the world. We have found that the protocol works basically in all of the different settings, irrespective of whether you are in Singapore, Salt Lake City, Boston, or wherever.

The other interesting thing is that we have deployed that down into pediatric and neonatal intensive care units and have found that, essentially, the same algorithm works. That was heresy for me as a pediatrician, who was always taught that kids are not little adults. In this particular instance, in fact they are probably little adults, or maybe adults are big kids.

That type of knowledge can then be rapidly incorporated. It can be aggregated very rapidly. The key point there is that while we can get to that target level of glucose and we can reduce the variability around it, this research will not answer the question about what is the best target to treat to. There has been some recent evidence showing that much tighter control of glucose in fact may not be the best thing to do in an intensive care setting. We may need to relax that.

This type of research may not help to answer that specific question, since we based it on best evidence of what people were saying was the best to treat to.

DR. SOX: I have been a lifelong advocate of computer-based decision support, until I got to Annals and started to have a sense for what evidence base that it works in looks like. It doesn't look really good. When I heard Marc's talk, I was totally dazzled.

I'm wondering, to try to answer your question of where do we go from here, how do we learn from the experience that you have had in a way that can be transmitted to other people in a way that they would find convincing for their setting. How, basically, do you get doctors to feel invested in decision support and want to pay attention to it?

DR. WILLIAMS: I think there are a couple of issues there. One is that Brent has established an advanced training program where he brings people in for a four-week course. Not only do you get the theory but you are actually required to bring a project to that course. The Health Care Delivery Institute works with you to help to have a success. There is that training aspect to understand the theory behind this and to also understand the theory of how to actually deploy it.

What that course has led to is development within other institutions of satellite courses that are either institution-specific or regional. In some cases, with the example of the Cystic Fibrosis Foundation, they said we think this is really important. We have huge variability in cystic fibrosis care. We are going to have everybody trained in this type of technique, and we are going to set up the measurement and collection system. If you want to be an accredited center, you must participate.

There have been a couple of excellent articles out of the CF Foundation that have shown some dramatic improvement in pulmonary and dietary management relating to this sharing.

One of the interesting things is that it creates an environment to share success. What you find is, when you begin to measure things, no one is the worst at everything, no one is the best at everything. There is variability. Some places that are worst in class are best in class in other areas. By sitting people together and talking about what works and what doesn't work, you can get a rapid learning environment. Then you also learn about what worked for deploying it and what didn't. That is a training perspective that I think has been, again, demonstrably successful.

The second issue relates to the barrier, I think, of publication. Frank Davidoff has published a couple of articles relating to the work that he has done looking at methodologies and organization of papers around quality improvement. I think those are beginning to define the landscape around how we should be presenting this information so that others can begin to learn from successful experiences around this type of improvement activity.

DR. SOX: It occurs to me that with computers you have the opportunity to randomize within an institution different methods for getting people's attention, for example. Maybe we need to get Steve out there to collaborate with you on some Bayesian studies that would generate some generalizable knowledge that would find a ready home at a journal like Annals of Internal Medicine.

DR. WILLIAMS: Yes, I would agree with that. I think that there are ways to do it. There was an example in pharmacogenomics where a children's hospital was offering a range of pharmacogenomic tests for inpatients that were going to be treated with medications. What they did was they had genotyped all of the patients. Then they actually looked at the medications that were used and assigned whether they thought it was a good match or a poor match based on the type of medication and dose. They found that there were significant differences in things like length of stay, restraints and holds, and adverse drug events.

They created a system which the physician could go into when they ordered medication. The system would say, this could be benefitted by a pharmacogenomic test. Do you want to do the test or not, yes/no. If you use that yes/no decision tree, you are now generating your prospective cohort. It is not in a randomized fashion, but you have a real-world trial where you can then measure your outcomes of interest, your length of stay, your restraints and holds, and your adverse events, based on did we follow the instruction or did we not follow the instruction.

I think that type of a process would lend itself to the type of analysis that Steve presented. I think that is a really intriguing idea.

DR. TEUTSCH: I think what you are describing is why culture and systems are so important. We often talk about research-based practice, where we get this data and then try to apply it, as opposed to practice-based research, which means that we actually learn from that system. Gwen was next.

MS. DARIEN: I actually just wanted to go back to Steve Goodman's presentation and comment. I think that one of the things that is critical to research is that people participate in it. I-SPY I know a lot about because one of my friends is leading the advocacy group on that.

The Bayesian approach is a design that appeals so much to advocates and patient advocates and those people that are actually going to go out there to help these trials accrue precisely because it is adaptive. I just want to reinforce the fact that research needs people. Participation in research, particularly cancer clinical trials, which I know the most about, is incredibly low.

The other thing that I think works about this trial is the collaboration across the different stakeholders. I would just make sure that we include that.

DR. SOX: I would be interested in your comments about how CER could be structured so that patients felt as if they were part of the game and that participation was an opportunity instead of something to be avoided.

MS. DARIEN: I think one of the most important things to patients and why the I-SPY trial and some of these adaptive trials work really well is what you were talking about in terms of asking the right questions. The right questions have to be questions that matter to patients and patient outcomes like quality of life. The questions have to be matched with their values. I think that is a really critical thing, and I think that is why the adaptive trials really appeal to people. They understand that you don't just go in with something that is fixed and you can adapt it as you are going along and as you are learning.

I think it is pretty horrifying to think that doctors aren't necessarily good lifetime learners because we want to think that they are.

I think the other aspect of the I-SPY trial that is a great precedent for other trials -- and I have been involved in a number of things -- is that all of the stakeholders have been involved from the very beginning.

If you want patients to buy into it, then you have to talk to the patients about it. You have to bring the patients into it from the beginning.

I-SPY1 had quite a number of MRIs and biopsies, but patients were brought in in the beginning to help design the decision-making tools and the education tools in order to communicate to the patients. That had an incredibly high retention level of people in the trials, and an incredibly high accrual rate.

I think that there are different ways of bringing people in in the beginning. I think that everybody wants to know the drugs or the protocols that they are given are effective. They also want to know that they are an improvement and that there is a learning process and that there is progress. I think that those are two ways to bring patients together, but I think there are, obviously, many more.

DR. TEUTSCH: Marc, let me get a couple of other people into the conversation for a minute.

DR. WILLIAMS: Even I have something very specific about patient involvement?

DR. TEUTSCH: Twenty seconds.

DR. WILLIAMS: The other place that I see, particularly specific to genomics, relates to a dilemma that has appeared about adverse events versus efficacy. I think that we have overly focused in pharmacogenomic research, particularly in the oncology realm, around the adverse events. If you take the UGT-1A1 EGAPP report, for example, there appears to be some evidence for increased efficacy, actually, in the patients that have the polymorphism.

If I were going to study that, I would be very interested and engaged in the patient set. What is more important to you, avoidance of these adverse events which are going to occur or eradication of the tumor. When I read that paper, I said, I would want a higher dose than the standard dose here because I'm willing to accept a higher adverse event rate. That is another place, I think, to engage.

DR. TEUTSCH: What are the important outcomes. What really matters.

DR. WILLIAMS: Then they can measure their own outcomes.

DR. TEUTSCH: Andrea.

DR. FERREIRA-GONZALEZ: You may want to take some follow-up questions because I have a different issue.

DR. TEUTSCH: We just have a few moments. What I would really like to do is get different issues on the table here. One of the options that we have going forward, having identified some of these salient things, each of which we could devote a long time to, is to figure out where we want to go next. We have issues here and I would like to hear what others are.

One of the things we hear a lot in this field is if you personalize things it is going to be hard to do it in a comparative effectiveness world. Is that an issue that we should be going down. There are issues of disparities that we know are important. How does that play out in genomics and in this field in general.

I would like to get some of those issues on the table here so we can figure out if there are some areas that we want to carry on. So it is fine to carry on with a different topic.

DR. FERREIRA-GONZALEZ: It is not so much of a question but a statement. As I continue to read about comparative effectiveness research and look at the different ideas that are being proposed on where you might be selecting different patients by genomic technologies and results of testing, we need to be cognitive that different technologies work differently.

One example could be of the clinical trials that you mentioned for breast cancer patients, the HER2-neu. If you do testing for HER2-neu identification by immunohistochemistry versus another method, you might have a different result.

I haven't heard anything, or read even, about the role or research needed on comparative effectiveness on some of these genomic technologies to be able to really focus on where you are going to be selecting the patient population.

With that also, as we continue to look at these types of studies where you are selecting patient populations or a group of individuals to go one or another route by using a diagnostic test or some genomic test, actually these tests should be done under the highest quality. Each test has a clinical validity and an analytical validity.

I haven't heard anybody talk about doing this in

CLIA-certified laboratories. As we go more into the genomics area, some of these tests might not really be available in a large number of CLIA-certified laboratories. This could be a very important issue, that we assure that the quality of the testing output to be selective in these different areas is of the best quality and done under certification.

The other issue that I was struck by is the amount of money that is being pushed and the need for infrastructure. Money is being given by NIH, AHRQ, and HHS to look at funding and comparative effectiveness research. We need to have coordination and maybe more transparency for the public on what different clinical trials are being used or what research is being used.

We could have a publicly available clearinghouse website of what is being funded and what are the results of what is being funded so we can come back and say this has already been done or this particular question has not been addressed. Maybe we could have something similar to the ClinicalTrials.gov website where that information can be assessed. I think this is a topic that I haven't heard discussed that I see as very important to this issue.

With the issue of the tissue banking, I think the quality of the specimen that is put in, not just the clinical information and the follow-up, is of huge importance. You might have the clinical information, but if the tissue is not appropriately stored or obtained, the data is going to be very skewed.

These are some issues of infrastructure that need to be dealt with or thought about before we dive into this type of comparative effectiveness research to make sure that the data we get out we can really rapidly translate into practice.

DR. GOODMAN: I'm determined to give space to Gurvaneet here to jump in, but I want to answer two things. First, I didn't go into nearly all the details of the I-SPY trial, but they are looking at exactly that issue of how HER2 is measured. They are measuring it three different ways, and they plan to shift from the immunohistochemistry model to the other technologies if they prove to be more predictive. That is embedded within it. They are doing that with several other biomarkers, as well. They are measuring them several different ways. That is part of the validation and improvement.

With regard to the tissue bank, I couldn't agree more. Everything I mentioned will require serious thought about how to create databases that will be usable 20 years later when they are called upon with new technologies.

DR. RANDHAWA: I just wanted to mention a few things. One, since we are talking about clearinghouses, there is always this challenge about information and quality improvement activities and how much of it gets published in peer-reviewed literature. AHRQ has started an innovations exchange clearinghouse for exactly that same purpose, so that we can share innovations and other people can learn from that. It just started taking in applications, and I can send you that link.

The second AHRQ activity is one we have funded more on the learning health care and practice-based research. It is the Distributed Research Network. There were two different models that we funded, but one of them is actually looking at how different primary care practices who want to benchmark how they are doing and compare each other and learn from each other can, independent of whatever EMR vendor and software they have, exchange that information. It can also be used for outcomes research.

The third part is the clinical decision support tools that I had mentioned before for BRCA testing. There will be an involvement of the patient in terms of getting the family history as well as shared decision-making with the provider. I think we will be learning something from that project.

DR. SOX: I wanted to seize on one aspect of your question, which was trying to achieve transparency as much as possible so that the public really understands what is happening. I'm in favor of that, except for one part that I'm a little worried about, and that is the research results themselves.

Steve, pay attention because I'm going to ask you a question.

Right now, I'm a strong advocate of journals and the processes that they go through to make sure that work is done according to good statistical practices and that the language that is used is transparent and isn't biased or slanted. Therefore, I wouldn't want to see research results published until they go through a process like that. I'm very old-fashioned.

Steve, what I'm wondering is whether there could be a time in which, with the appropriate design of research, perhaps particularly adaptive trials, we could skip the journal part. In other words, things would be done in such a way that the role for journals would be reduced perhaps to editing reviews of subjects like that. Do you think there will always be a call for journals?

Steve, by the way, is the editor of Controlled Clinical Trials, so he is an expert on this.

DR. GOODMAN: Clinical Trials. Controlled Clinical Trials doesn't exist anymore.

I think we are always going to need impartial arbiters of the science. My favorite quote on this was from Jan van den Broek, who gave a talk at I think it was the 50th anniversary of The Lancet. He said, this fantasy that we could have results just poured onto the Internet is just that. If we started doing that, I think the quote was -- and this wasn't from Yogi Berra or Mark Twain -- something about how an enterprising band of young scientists would get together to vet the research and organize and deliver it to scientists so that the journal system would be immediately reformed.

I think that the independent oversight and review of research will have to be retained. I think researchers themselves, both because of training and because of inherent intellectual conflicts of interest, aren't always the best judges of their own work. I would just leave it at that.

DR. FERREIRA-GONZALEZ: I don't think I was talking about putting all the results but what is being done. It is also important to realize that some things are not published. Negative findings are not very publishable, but they are extremely powerful so we don't go down the same road. How do we deal with that?

DR. GOODMAN: As you said, a lot of this is being done in ClinicalTrials.gov. There are also several international efforts by WHO and some others devoted to developing standards for reporting results of research. It is starting at the RCT stage because those are the most structured ways we have of doing and reporting experiments. To what extent this can be extended to all research or other forms of research I think is a really complex technical challenge. Even with RCTs, it is very, very difficult to know what do you put out there, what do you put out there vetted or not vetted. Do you put analyzed data.

A lot of groups are struggling with this, but it is very much a subject of international collaboration and activity as we speak.

DR. SOX: Just knowing that the research exists, that somebody tried, can help a lot.

DR. GOODMAN: Right. That is what ClinicalTrials.gov does, at least in the clinical trials domain.

DR. SOX: You don't have to have the results to know a lot more about what the body of evidence might look like if every trial was in it was registered.

DR. FERREIRA-GONZALEZ: There are also the negative findings that don't make it to peer-reviewed literature. I don't have an answer to this. It is very important that investigators know, and I don't have an answer how to do this.

DR. GOODMAN: That is what clinical trial registration does.

DR. FERREIRA-GONZALEZ: The registration, but there are no results or anything of the negative.

DR. GOODMAN: That is the beginning of being able to go back to the company or the investigator. You know what the denominator is. The results may or may not be there, but in theory you can go ask them.

DR. TEUTSCH: Some should be on ClinicalTrials.gov, which doesn't include genetic tests, as far as I know.

DR. GOODMAN: Right. No, it doesn't.

MS. WALCOFF: Thank you all. I think this has been a great discussion. I just want to bring it back to Gurvaneet and Steve. In both of your slide sets you included a page on solutions and future steps. I actually thought it might be very helpful for this group to talk about that for a moment. If each of you could suggest one thing that the Department of Health and Human Services could do in this area to forward these approaches, whether it is eliminating regulatory barriers at FDA, or consolidating or linking up the innovation clearinghouse that Gurvaneet talked about with the other databases that all the different departments and agencies are putting together, what would those things be?

DR. RANDHAWA: You go first.

MS. WALCOFF: You have sort of answered the question already.

DR. RANDHAWA: The only thing that I'm struggling with is what is my top priority. There are so many of them that are competing. I think the fundamental issue is an infrastructure that can get at what is happening in health care and learn from it. That would include informatics as well as better clinical data collection while maintaining the privacy and confidentiality of patient information. That would be my Priority No. 1.

DR. GOODMAN: I don't know that I actually have anything to add over what I said. I guess it is two pieces. One is to create sources databases from past experimentation that allow us to test current hypotheses as quickly as we can and to reserve the prospective component only for those questions that absolutely can't be answered to a sufficient extent with adequate past data.

We already don't have adequate past data, so we have to look forward and start creating our past in real time. In terms of HHS, I think what I mentioned before is thinking about how to formally support the increasing longevity of the data that we gather in the context of clinical research, with RCTs being the natural first place because it is the highest quality and the most structured. That, in a sense, offers the biggest bang for the least buck.

Secondly, going forward, focusing on the resources, which include development of informatics pieces, the tissue storage, long-term follow-up, everything that is involved in using methodologies that will get us answers a bit more quickly and more efficiently. As I said, right now everything is built almost from scratch for each trial. We need to increase the resources available for the development of the software, the training, the informatics backbone, all these things. I guess that would be how I would summarize it.

DR. SOX: If I had the Secretary's ear, I would be urging her to make a really serious effort at coordinating CER across the different agencies of HHS, as well as the VA and the Department of Defense, so that outcome measures are standardized using the instruments that are widely available and validated, so that as much as possible we end up with research that can be compared even though the funding agency may be a different one.

In addition, as much as possible, promoting collaboration between agencies and funding research on high-priority questions and conditions. A serious effort at coordination.

DR. GOODMAN: One footnote to that is -- and this is something that is to some extent a priority already, I think -- doing everything they can to enable the extension of this research into community research networks. Most patients are not seen in academic centers. This is being done, again, piecemeal on a disease-by-disease basis, but we have to bring in the community practitioners if we are going to do CER or, really, almost any research that requires substantial numbers and answers practical questions faced by doctors where they basically currently are.

DR. TEUTSCH: This has been a great discussion. We are now at the point where we have to figure out what we are going to do from here. Is there a role for all of us. I have heard a lot of good issues.

We had early discussions that said we need the evidence before we can actually move some of these genetic tests forward into practice, so it seems like this is a critical issue for us. I have heard issues that are surrounding what are the studies and the study designs, how do we encourage that, how do we build the right infrastructure, whether it is laboratories or biobanks or standards and metrics. I have heard, how do we move into a learning health care system, how do we engage patients and consumers into doing things that matter to them so that we can build this enterprise.

There are some issues that relate to the fact that what we are talking about is a very rapidly moving field and the issues of personalization. There is a general sense I think I hear occasionally, not that I buy into it, that there is some dichotomy between personalized health care and the information that comes out of comparative studies because they are more population level. I think we can make those things work.

I think there are a bunch of issues here. The question is, are there some of these that we are well positioned to take on and work through. The proposal I would like to put on the table for you to consider, since I doubt we will be able to get to anything like a scope of work right now, is that we actually form a small group to sort through the issues and bring back to us next time a distilled and considered list of things that we could do and some recommendations about whether we should go forward with some of this work and some ideas on the scope of work.

DR. WILLIAMS: If I could add two things to the list that may also help to focus this. I look at this, we clearly have to look at it through the lens of genetics, health, and society. At the present time, we don't know what the IOM report is going to look like and what their prioritization is going to be. That will be forthcoming.

The second issue is that we will presumably have the round of funding announcements from the first round of the Recovery Act grants. I think that is in September that that comes out. I don't know whether it is even possible -- Michael could probably answer if he were here, or maybe Alan can -- whether we could somehow get a list of at least the general pots of funding to see what we might consider to be in the genetics and personalized medicine realm that was actually funded in the first round.

That would also give us an idea to say are there priority areas that we have identified previously that in fact somehow escaped being funded in this first round. That could also help to formulate where we are going.

Those would be the two things I would add to the list.

DR. TEUTSCH: I didn't mean to make this a fixed list, either. Having heard this discussion, knowing the documents that we have done previously, I think the group could tease out whether there is an agenda for us.

DR. GOODMAN: Could I make one comment? The word caBIG wasn't mentioned here, but the experience there is interesting. Of course, caBIG stands for Cancer Bioinformatics Grid. It was a monstrously ambitious effort which I think everybody is fond of deriding, but it has made, although much slower than I think they envisioned, real progress. Where the progress is, is not in the tools themselves but in the standards, in getting people to talk to each other, which relates to what Hal said.

You could think on your agenda of what standards there could be in the domain of genetic testing that would enable both sharing of information and establishment of quality standards. I wouldn't even know where to start.

I don't think that is where they thought the bang was going to be. They thought it was going to be in all the bioinformatics tools. Ultimately, many people are building their own tools but to those standards. It is like the iPhone model. They unleased this huge marketplace appeal. People are building applications using a common set of standards.

I don't know that anybody can dictate what those tools could or should be, but if you have a set of standards that they have to meet, you will move things forward.

MS. WALCOFF: I was just going to say on dictating that that is one thing the federal government can do with respect to federal money. If you are going to give out grants in this area, you can dictate the standard, whatever it may be, be applied across the board for all such grants, whether it is how you file your information or what have you.

DR. TEUTSCH: If folks are okay with that, I would like to see some volunteers who could help pull all of these threads together and help us shape and bring back something.

MS. DARIEN: Do you need a patient voice?

DR. TEUTSCH: I would be delighted to have a patient voice, Gwen. Dr. Williams. Andrea brings a laboratory perspective. I think we will probably want to bring some of our federal partners into that.

DR. WILLIAMS: I would think, at the very minimum, Gurvaneet and someone from NIH.

DR. TEUTSCH: Can we start with that core group? Certainly Alberto, and Liz, since she is sitting here.

If there are others just let us know. Marc, I know you have given a huge amount of time. Are you willing to help lead this enterprise? That would be great.

Let me again thank our terrific panelists, who are a superb group of folks.

[Applause.]

DR. TEUTSCH: We appreciate all the insights, direction, and leadership that you all provide. Hal, all the best with whatever comes next.