A National Web Conference on Evaluating Measures of Success Using Clinical Decision Support
(January 21, 2009)

Good afternoon, ladies and gentlemen, this is Brian Dixon from the AHRQ National Resource Center for Health Information Technology. In a moment I will turn it over to Rebecca Roper of AHRQ and she will introduce our panel and start our conference today. First, I want to go over a couple of logistics items related to the WebEx environment. First of all, you may have noticed when you logged in that the toll-free number may not have appeared on your screen for the audio portion. The correct number is 1-866-699-3239 and then you will enter a meeting number 669612533, if you would prefer to listen on the phone rather than from your computer speakers. Also, the format today we will have presentations from each of our panelists followed by a short question and answer period for that individual speaker and then we will have extra time at the end for general questions and answers. So if your question does not get answered the first time that you ask it, we will try to get it in the general Q&A at the end of the session. Also today I will begin a quick poll that will appear on the right-hand side of your screen. These are questions designed to get additional information such as suggested topics you'd like to see in the future. This information is important to us at the National Resource Center, so, I hope that you will submit your thoughts to that. When you log off today you will receive a survey in a pop-up window asking for feedback on today's session and that is important as well and helps us to maintain the quality of these conference calls going forward.

With that I will turn it over to Rebecca Roper of AHRQ.

Good afternoon, this is Rebecca Roper and I am joined by Dr. Zayas-Caban and we welcome our four presenters this afternoon, for our fourth in the series with respect to clinical decision support. Today we will be focusing on evaluating measures of success using clinical decision support. As Brian articulated earlier, we are going to use a little bit of a different format. We will have three distinct presentations about 15 to 20 minutes in length. They're full of wonderful, rich material. At the close of the presentation, we will have a 5 minute opportunity for folks to submit questions to the panelist and we will be grouping those questions together and sharing them with the presenter so that everyone may hear the responses of our presenter. And then as time allows, after all presentations have been made and short question and answers have occurred, we will open it up for additional questions.

Our first speaker we're pleased to let you know is Dr. Charles Friedman. He is the Deputy National Coordinator for Health Information Technology in the office of the Secretary for Health and Human Services. He leads the office of the National Coordinator for Health IT efforts related to clinical decision support and is the author of a textbook on evaluation methods and biomedical informatics. After that, Jerry Osheroff, Chief Clinical Informatics Officer for Thomson Reuters will present. He is lead author of the 2006 white paper commissioned by HHS entitled "A Roadmap for National Action on Clinical Decision Support" and is lead author of a series of popular guidebooks for CDS implementers. And together, Drs. Randall Cebul and Peter Greco will give a presentation. Dr. Greco is a professor of medicine and epidemiology and biostatistics at Case Western Reserve University (CWRU) School of Medicine and Director of the CWRU-Metro Health System Center for Health Care Research and Policy. He will be joined by Dr. Peter Greco an assistant professor of medicine at Case Western Reserve University of Medicine and consultant in EMR implementation for the Metro Health System.

With that, I will welcome Charles Friedman.

Thank you, I hope everybody can hear me.

My talk is a bit iconoclastic and uses a baseball metaphor to make an important point about evaluation and informatics in general and one that I think that is particularly relevant to evaluation relating to clinical decision support. I modified this talk to relate specifically to clinical decision support from an earlier talk that I gave using the same metaphor back in 2004. If you would like to refer to that paper, you see a reference to it hear from the Journal of the Medical Library Association.

You may be asking what evaluation has to do with baseball and this is the basis of the metaphor. It turns out for those of you who may not be familiar with this, there are two basic approaches to scoring runs in baseball. One which is called powerball where you basically try to hit the ball over the fence every time and every batter is viewed as a potential home run. The second strategy, which is called smallball where you play for one runner at a time and somebody may get a single or walk and then you bunt that person to second base and hope that player steals third and then scores on the sacrifice fly. The metaphor will play itself out over time but I am going to argue, relating this back to evaluation that we need to play more smallball than powerball when doing evaluations in informatics in general and specifically in evaluations relating to clinical decision support.

So now let's look at powerball as that metaphor applies to evaluation. If you're playing for the evaluation home run, so to speak, the way that you evaluate any project is by doing one big study. And there is really only one question of interest and that would be whether patients or the population are healthier or, if it is a study of a research application, whether scientists are more productive because of this intervention when all is said or done. Basically, if you do powerball evaluation, the only method you would want to use is the closest approximation that you can come to a randomized trial. Powerball approaches dictate that no evaluation is necessary until the project is over and really the only result that you are interested in is if there is some difference between the groups in your randomized study.

Smallball evaluation, as the metaphor would imply, is a step-by-step evaluation. Instead of one big study it consists of small studies. And through this smallball view of evaluation you view each stage of a project lifecycle as one that presents important needs and new and important challenges for evaluation. At each phase there are many questions of interest and a comprehensive evaluation would comprise many small studies instead of one big one.

Here you see in this slide the points played out in a little more detail. Projected against what I think is a reasonable depiction of a project lifecycle before an information resource is deployed. During the project itself, when the resource is deployed, you might be interested in some behavior change on the part of health care professionals and then you might be interested in some behavior change on the part of the clients of those professionals. And that will lead to some outcome. Against this backdrop of a general schematic of a project laid out in time, powerball evaluation basically skips over from the before to the after. By contrast, smallball evaluation would look at questions that arise at each stage of the project lifecycle, for example a study of what the exact needs are for projecting this into a specific clinical decision supports system, whether the design of the system addresses the needs when it is still a design.

After initial deployment, it might look at, in the clinical arena, what is the buzz about the thing soon after its deployed? You might look at who used it and for what purposes? These are smallball studies looking at smallball questions that are very important. Looking at behavior change in the smallball sense, the question you might investigate is whether behavior is correlated with use. In other words, did the patients of the providers who use the system more realize greater effects? This is the kind of study that would not require a control group.

So this is a picture of how these two different approaches to evaluation might apply against the backdrop of a general time-based schematic of any intervention. So there are clearly arguments to do things either way. The argument for powerball has many components to it. First and foremost, perhaps this is what people expect and, maybe even more first and foremost, it can be seen as what peer review committees expect. Powerball has the virtue of using the message of evidence-based practice and the results of powerball studies generate concise results, like an effect size. And the results in principle, and this is very important as a strength of powerball, each study can be meta-analyzed to generate results with greater generalizability. Finally, in some domains powerball studies are the only ones that can get published.

So the argument against powerball: it is expensive, slow, you don't really learn anything until the intervention is completed and has some time to play itself out into stable effects. Powerball requires freezing the intervention and this is a big issue in informatics because the developers of clinical decision support systems, as they learn more about how they are functioning, want to change them and improve them during deployment. And often powerball studies prevent that. It requires intricate controlling of the environment and there are also a lot of questions some of which I illustrated before, that it can't address.

So what are the arguments for smallball? It can be done on the cheap, so one can argue that it's always possible. It's also agile, it attaches itself to the lifecycle of a project and enables design and implementation to become self-correcting processes. You can learn from each smallball study and apply the results of it to the next stage of the project. It does not require freezing the project or controls, and as it so often happens, if the system does not evolve in the way that was originally anticipated, you can study how it was actually implemented as opposed to what was envisioned at the outset.

As in baseball, for those of you familiar with the low-budget baseball teams that tend to play smallball, smallball evaluation is best matched with low-budget operations, and it might be the case that everything in 2009, given the state of the economy, is a low-budget operation.

Now, applying this directly to clinical decision support. Some of the reason why a smallball study may be of particular value as they apply to interventions where the focus is clinical decision support. Prior to the deployment, smallball studies can, in terms of what you learn from them, narrow the broad cultural gulf between the end users of the system and the information professionals who build these resources. It is very often the case that the builders don't have the insight on what it is like to be a health-care provider unless studies that provide that kind of feedback to them are built into the development process. Smallball evaluations can also bring the real needs of the clinicians where the system is being deployed into clearer focus and address, I think you'll hear about this later, the important issue of how a deployed clinical decision support system can be optimally fit into a professional workflow. During early deployment and testing, several smallball issues arise here relating to clinical decision support. I think it is clear in clinical decision support if you look at a chain of events from the issuing of advice to the point when that advice might have some benefit on care, a lot of things have to go just right in order for that benefit to occur. And smallball evaluations can show if things are not working out as hoped and they can show exactly where the chain is breaking down, so appropriate remedial activities to the system can be taken. Smallball evaluations can maybe not show, but certainly point to the distinct possibility that a system in its early stages may be doing harm, and that has been a concern from the outset with clinical decision support systems, especially in early deployment.

After deployment, obviously smallball studies cannot demonstrate the magnitude of an effect with the scientific internal validity and precision that a randomized controlled study can. It is often the case that given budget and other constraints, smallball studies are the best that can be done. This is for a variety of reasons including the complexities of patient care settings that preclude other things that you need to do to implement a controlled trial. Smallball techniques like dose effect, like I explained earlier, or extent of use studies, extent of compliance studies, as you will hear about later, could be extremely informative even though they don't hit the home run. And finally, every informatics project that I have ever been a part of has unforeseen outcomes and unanticipated effects and a smallball study might be better than their counterparts in detecting those.

I am showing you this at the end. It really was in many ways the motivation for this talk and my coming up with the smallball/powerball metaphor. This is an excerpt from a summary statement received by an applicant for a grant requesting $100,000 over two years. And, as you can see, this project's scope and budget could not possibly support a powerball evaluation. If you read the bullets here it was severely criticized and ultimately not funded because it did not have a powerball evaluation. So I think there is a lesson here about how to right size evaluations and match what you do and how you do it to the size of the project.

So bringing this to conclusion, probably the most important bullet in terms of philosophy that I am going to offer here is the one on the top that suggests my own view. It is better to develop some insight into something really important than it is to find nothing in pursuit of knowing everything. This is one way of formulating the argument for smallball. We live in a practical world and it follows that the evaluations you should do are limited to the evaluations that you can do and as I said earlier, very often the resources simply aren't available to do powerball evaluations.

I want to emphasize that smallball evaluations can be done rigorously. Just because they are smaller and don't employ the methods of randomized trials does not mean that they are sloppy. By no means am I making an argument for sloppy evaluations. This last slide just summarizes the implications for evaluation going forward. If you buy into this distinction, clearly we need some powerball studies but all studies need an evaluation, which makes the case for smallball evaluations. Maybe it means that smallball will be the rule and powerball is the exception. Smallball gives us the ability to make evaluation more agile and I think the real pathology which motivates my talk is an expectation, as you saw in that summary statement excerpt, that every project will have a powerball evaluation or, as a sort of corollary to that, every project will have a powerball evaluation or nothing.

For those of you who are baseball fans here is a tribute to the 1963 Los Angeles Dodgers that beat the New York Yankees, the ultimate powerball team, in four games by playing smallball. They are the virtuosos of smallball and hats off to them.

Thank you, Dr. Friedman.

I just wanted to echo this at AHRQ, as many of you are aware, in the fall we publish three new health IT funding opportunity announcements that provides a continuum for grant support. Although we did not codify the concept of smallball as elegantly or concisely as Dr. Freedman, when you look at the requirements with respect to articulating an analytical plan and we specify that the appropriateness of the analytical plan needs to be well-matched to the study designed and discuss qualitative and quantitative measures and as Dr. Friedman said, the scope and funding can often preclude these powerball approaches. So I just want to echo that for these funding opportunity announcements that we expect to

see a lot of ball playing at varying levels. I guess it's a coincidence that we're starting off here at spring training, at least in terms of new health IT grants coming in.

Let's see what questions we have for the panelists.

One particular question is how do you get funding to do smallball evaluations?

I would see, as I said, the three health IT announcements in particular the R03, which provides 100,000 up to two years and R21, which provides 300,000 total cost support for up to two years. And, depending on the different components from the specific aims, the more robust evaluations of an R18 may be of such complexity in research design that there may be small ball elements.

If anybody has any more questions, please feel free to send it to all panelists.

Okay, we will let questions percolate. We will now take the opportunity to hear from Dr. Jerry Osheroff.

Thank you, Rebecca.

What I would like to do next is build on the important and practical considerations that Dr. Friedman laid out for us. Looking at things in a little bit more detail through the eyes of people in the trenches who are trying to improve outcomes with clinical decision support and are faced with the very basic questions of did we actually do good or do or did we do harm?

So I would like to throw out some common challenges related to CDS in general and implementing CDS, particularly what we hear a lot from implementers and then share some pearls that maybe helpful to address those challenges based on some work to synthesize best practices for implementing clinical decision support that I will say more about in a moment.

In this latest guide book there is an entire chapter devoted to evaluations, and we will dive a little deeper into that as well as address overcoming specific challenges. And then the goal that we would like to shoot for is to have everybody come away from the series of talks with something very practical you can take away for your own CDS implementation and evaluation efforts. We hope you will be thinking of questions pertaining specifically to your activities and share them with us.

Clinical decision support, as we all know, it's not an end unto itself, it’s the means to an end. The backdrop, especially these days for our CDS efforts are the major drivers that are happening out in the healthcare environment. These drives are leading to very powerful performance improvement relating to quality, safety, efficiency, cost etc.

And here are some examples of those challenges. We know that payers have recently begun not paying for adverse events that shouldn’t be happening. There has been a strong move toward pay-for-performance and that is only increasing as we go forward in the value-based healthcare world and we're starting to see this filter down to hospital and healthcare delivery. So, individual employees’ payments are increasing depending on meeting performance targets.

There is a lot of talk around transparency and accountability and these are some examples of the forces that provide the backdrop for our CDS effort. So, well beyond the early days when we were experimenting with making sure that people on Digoxin have good potassium levels, now there is a much

broader sweep of major drivers that we are trying to deal with. And organizations that are spending large amounts of money deploying health IT systems are getting help from things like the Leapfrog CPOE test that are even more specific drivers that are coming from the payer side of the world that are helping to ask the question, are CDS interventions where we want them to be?

So I think all of the folks on the phone would be excited about and appreciate the notion that clinical decision support will be a powerful suite of tools to address these major challenges that we outlined on the last slide. I think an important cautionary note comes from the study from 2005 and others like it, this is entitled High Rates of Adverse Drug Events in a Highly Computerized Hospital. I think the warning to note on this study is that this is an organization that was seemingly doing all the right things. They had very smart people, they had health IT systems and clinical decision support and after they deployed those things the major kind of adverse outcomes that they were trying to avoid were found to be persistent. This points out that a CDS system is not just some sort of database that you buy or some lines you check off on your health IT implementation plan. It requires a tremendous amount of thought and planning to avoid the results reflected in the title of this study.

So from the CDS implementation and evaluation perspective, some of the major challenges that we see as organizations, we have to question, how can we get the resources and attention for our CDS evaluation effort. So we deployed our CDS system and we are not sure exactly what kind of effects it's been having and for that matter we're not completely sure what we deployed because we are in such a frenzy.

Another common sort of concern is why aren't clinician's responding well? Why are there a lot of overrides? Why aren't they using our order sets? And the broader issue of, our information systems are not doing all of the things we want them to do in terms of good CDS, so what ways are best to deal with those constraints?

These are major challenges that are faced by many different organizations and there has been a tendency, such as the individual on this slide, to deal with very hard issues in utter isolation. The good news is in the past couple years there has been an effort to bring folks together so we can synthesize and distill what we need to do from both a national perspective to make CDS more efficient and effective, nationally and then also from a provider's perspective. What sort of things do individuals and organizations need to do to get their own ship right in terms of leveraging CDS?

This slide has a couple of examples, one is the Road Map for National Action on Clinical Decision Support, which calls for the development and dissemination of CDS best practices and also for the intensified efforts to sort of get the evaluation things right so we can understand what is working and what is not working so we can do that more broadly.

On the bottom part of the slide there are two examples of efforts to pull together the expertise that has been learned in isolation for many of the leading organizations that have been working on clinical decision support to pull it together in such a way that it is available more widely so we can move faster and farther to where it is we're trying to go.

The book on the right is a highly collaborative project that will actually be published within the next month or so. There are nearly 100 contributors to this project that will be posted on the AHRQ website. And some of the points here that I will be calling into more detail in the next slides are the ongoing conversations on how to bring people together to solve these challenging problems.

The next couple of slides I will pull out some material from the latest guide that will be useful to you in your CDS implementation and evaluation efforts.

This provides the overall outline. It follows a typical performance improvement process of identifying the governance issues together and figuring out what the goals of your projects are, looking at workflows etc., and then, in the lower left hand corner, you can see assessing what it is that you've done and including that as part of a performance improvement loop is a critical part of the whole process.

So in this guide and in the roadmap, we start with a very broad definition of clinical decision support, which is essentially to provide clinicians and patients with data which enhances clinical knowledge and supports patient care. And this makes clinical decision support not something that an organization can say "well, we'll get to that when we get to CPOE or other advanced clinical systems" but it also includes other sorts of things happening every day in hospitals all over the country and the world. And going back to the smallball evaluations, answering the question of what exactly are we doing today and what benefits are we getting from that are incorporated into a very broad definition like this.

Again, in these areas in the roadmap and the guides, we try to be specific about what the kinds of CDS intervention types are because in both the literature and in the actual implementation there tends to be a tremendous focus on the alerts and reminders. They are certainly a powerful type of CDS intervention, but as you can see in this slide, there are many others, things that help with relevant data, creation of orders, reference information that can answer clinical questions and all of these things are very important in the kind of tools in the clinical decision support portfolio.

So part of the essence that we propose is a success formula for getting it right, so if you are going to improve specific care outcomes and these are the things that are the focal points for some of the drivers that I mentioned earlier, it sort of boils down to these five rights. The hypothesis is if you can get the right information to the right stakeholder in the right format, again leveraging broadly the kinds of CDS interventions that are available and looking broadly, looking at patients as well. If you can deliver that through the right channel and at the right point in the work flow, then by getting these five rights right, then that becomes the success formula for achieving this type of formula which is becoming increasingly essential and imperative.

When you look at medication management specifically, which is the focus of this new guide I mentioned, you can see basically the workflow steps and the medication management cycle around the loop and some of the objectives. Each link in the chain must be strong to insure that a good outcome from medicine use that we all seek will in fact be delivered.

This is way too complex to go over in a brief talk like this, but this is an example of what happens when you take those five rights and lay them against the specific goals that you are trying to accomplish at each one of these steps and sort of answer the questions of what does the configuration of the five rights look like so we can accomplish these specific things with medication management or some other objective we are trying to reach.

The point here is that we can use frameworks or models like this to think very broadly on what we're trying to do and how to do it, which of course, has implications for the evaluation as well. In each chapter

of this guidebook, each chapter begins with a series of key tasks and implementations. I would like to run through them quickly because each has significant implications for the evaluation process.

In the first chapter, having to do with CDS, the basics to go over are some of the material I just mentioned about thinking very broadly about clinical decision support and what you are trying to do from a systematic, five rights perspective.

The second chapter, which deals with establishing your foundation and some of the pearls that we identified, are making sure you have engaged all pertinent stakeholders and you have appropriate governance in place for your CDS activity. You have to make sure to prioritize your priorities and targets carefully and make sure that they are aligned with organizational imperatives. Once this is done you can see where you are with specific objectives.

And workflow examination is really critical, not just guessing or imagining or having a bunch of people from the informatics team off in a room, really looking at the real world, just a walk through can do wonders- really building on the smallball evaluation theme, just a walk around can do wonders to understand what people are doing or what they think they are doing or should be doing.

And looking at things from the perspective of each individual system you are already deploying, there is a lot of health IT running around all different types of health organizations. And we emphasize the point that many or most of them offer the opportunity for clinical decision support and that should be filtered into the mix.

And the notion of trying to figure out once you have identified what the specific targets are, applying the workflow analysis and five rights on how to deploy the intervention to achieve the objective.

If you take that in a broad perspective, that is one of the antidotes to getting around this tremendous and ubiquitous alert fatigue problem.

And when you go ahead to deploy the intervention, if you look at a lot of the heroically failed CDS implementations, a lot of it derives from the perspective that we will do clinical decision support to the recipient as opposed to it being a shared process, which hopefully some of the other pearls have emphasized, that this is something that you do with the constituents. Getting the stakeholders to have some sort of a shared vision of what the problems are and what it will take to fix them rather than doing CDS in a confrontational kind of way, it becomes a shared activity toward achieving joint goals with the various individuals, organizations and constituencies.

And then there is an entire chapter devoted to measuring the effects of what it is you have done and attempting to refine the program. And some of the pearls here are just doing it. Going back to smallball evaluation, they figure if they do any kind of evaluation it will be a massive undertaking and they don't have the resources to do that. So they don't do anything or nearly enough, so just doing something. Again, I think Dr. Friedman has given us a nice way to think about it so we can proceed and then linking what it is you are assessing to the organizational priorities and then finally managing the knowledge assets and the processes both proactively and systematically.

So, diving deeper into the evaluation chapter, here's a little bit more of the key tasks and key lessons. Again, as with anything in this business, a systematic and proactive approach can be very helpful and powerful.

So specifically with the measurement activities, what some of the key measures are, how you are going to measure and why. We go into a lot of detail to look at structure, process and outcomes metrics, I'll go into that more in the next slide, and Dr. Friedman talks about that a little bit, looking at both intended effects and unintended consequences.

And then using this as a part of the continuous evaluation process, even the evaluation isn't an end unto itself. It is meant to draw you closer and closer to the priority targets and imperatives that you are trying to address. And some of the key lessons, again coming back to the issue of where the money is going to come from, in a care delivery organization if you have your activities in such a way that their priorities are closely aligned, particularly with the executives being compensated on how the specific targets are achieved, that can create a significant amount of organizational wind behind both the CDS and CDS evaluation. Another pearl is to have a rich and complete baseline for the things that are the highest priority targets and systematically using elegant CDS and evaluation to demonstrate that you are, in fact, getting closer and closer to those imperatives that are driving the organization. Rather than, as many organizations have done to put all of their eggs in the CDS deployment basket, putting themselves in a situation where there are not adequate resources to do the evaluation, through a systematic approach tackle this stuff right from the very beginning of the process.

So a colleague of mine at Thompson Reuters who is a Senior VP of customer experience shared this acronym with me. It is METRIC; he says that the approach you should use is measure everything that really impacts customers. And that would include patients, clinicians, the organization itself and many other stakeholders. So these things that really impact customers can be broken down into three categories. Structure measures--what exactly do we have deployed and how it is configured. And then the process measures--how is what we are doing affecting end users, their decision, actions etc. Who is using what we have deployed? Are they finding it useful? And then the outcome measures--are they bringing us closer to where it is we're trying to go? Looking at things like safety, again like the Leapfrog test and ADE triggers, cost goals, quality goals, satisfaction, etc.

So again, when you look at this, and I will emphasize some of the points that I was making earlier, it can be very daunting to figure out where to start. But to the extent that the organization's clinical decision support activities are tied to other performance improvement initiatives, which are basically ubiquitous to care delivery organizations, that is the extent to which you can leverage that infrastructure and resources and attention and interest. All of those can be very powerful allies in getting the CDS activities done rather than looking at the CDS implementation and evaluation in a compartmentalized sort of way.

So hopefully some of these comments I've made in the past moments give you a hint to what some of these answers, as we propose them, might look like when facing these major challenges. In other words, issues of how do we get resources as far as CDS evaluation and how do we make sure that clinicians respond effectively and appropriately to what it is you're trying to support them in doing and figuring out how to leverage, as best you can, whatever level of clinical information system that you have deployed.

So some of the things like the governance issue, make sure it's tied to organizational priorities. Not approaching CDS as something that you are going to unveil and pull off a sheet and then say "tada" and hope that clinicians respond appropriately. Instead, have it be a process from the beginning that you're doing with the recipient and taking a very systematic and five rights oriented approach so that you know what information you're delivering to whom and what it's accomplishing.

I want to mention very briefly some follow-on efforts I noted earlier. The preface to this new book is entitled, "this is not a book." This is to say that this is more of a conversation, the goal of which is to bring more and more implementers, including the 200 and some odd participants on this call into the process of learning from each other and teaching each other. We are in the process of building a Wiki to underpin this ongoing conversation about strategies about clinical decision support to improve these high priority outcomes. There is a task force run by HIMSS and the Scottsdale Institute that is building on this work with a more specific topic, in this case looking at dena stromboli embolisms. And looking at very specific strategies for deploying clinical decision support, building on all of the notions that I described earlier to optimally advance outcomes in this area and then to scale that to other topics. And many organizations, some that are listed here, are having many conversations about moving the ball forward. Many presentations and question and answers such as the one coming up shortly are other opportunities to broaden the conversation and collaboration.

So, just some things for you to think about as we move forward into the other presentations and the Q&A regarding how this might apply to your organization. You might think of some performance imperative that your organization is facing, are the objectives being accomplished, are they being fully leveraged, etc. And from the collaborative perspective, ask how we might get engaged in some of these broader efforts?

On this slide I have some of the references that I mentioned earlier. The material on this slide has my e-mail on it and I would welcome hearing from anyone after this.

Thank you and I will turn things back to Rebecca Roper for the Q&A.

Thank you very much, Dr. Osheroff. We received a question with respect to this being the fourth in the series, how does one access the previous webinars? If you go to the NRC website at healthit.AHRQ.gov and go to “past events” they are in chronological order, beginning on September 19th when the “using CDS to make patient centered care decisions” webinar was presented; you will find both the recording of the presentation, the PowerPoint presentation, and a transcript.

If you go to October 27, 2008 you will see the CDS discussion on workflow issues and the third on November 18th, which is “how is clinical decision support used to monitor and improve population health?” And in a couple of weeks the PowerPoint presentation we are using today will be made available in a similar fashion as well as the transcript. We will have available one of Dr. Osheroff's chapters and have a link to that chapter, which he discussed in his presentation.

Rebecca, there is a question here that I would like to respond to. Is there time for that now?

Go ahead.

The question is, since CDS is very broad and it includes information for non-algorithmic decision-making and given that primary care practices are so broad, how do we know if CDS helps clinicians with the decisions they need to make? I think that builds on both the comments that I made and that Dr. Friedman made. One of the ways to look at that and there have been some studies, for example, Partners HealthCare where they use info buttons so you can click on a link next to a drug name or a disease name to get more information about that drug or disease. So evaluation studies have asked about when you click on these links, primary care providers, nurses and others, is the information you are receiving changing your decision and is it useful? And the punch line of those studies is that that particular intervention does, in fact, turn out to be useful in that way. So if you think about what we're doing as

trying to support decisions rather than make a decision, then assessing the extent to which the support that we are providing is useful to the people that we are providing it to, and doing so in a systematic and peer-reviewed and published way, I think those are good examples of how we can ask the question and the data that we have so far provides information that, in fact, it is useful.

Okay. I have another question that I would like to pose to Dr. Osheroff and then the other panelists may care to respond as well.

It is great that there are organizations like Leapfrog that have CDS standards that hospitals are expected to meet, unfortunately many requirements cannot be met with existing vendor offerings so that one, creating excessive alerting and systems, and or two, requiring excessive resource allocation due to complex coding rules. These primarily arise as a result of limitations and inconsistencies between clinical knowledge bases and functionality shortcomings of vendor software. How do you think federal or state organizations can help hospitals pressure these organizations to improve their products?

Dr. Osheroff would you like to respond?

Let me take a stab at this from a couple different perspectives. One is that the Leapfrog folks are interested in helping move the ball forward so I would direct the person who asked that question to the Leapfrog CPOE consortium because they're trying to get a bunch of different stakeholders together to move this forward rather than having an impassable punitive test. They are very interested in using the work that I have mentioned here to move things forward. I think that is one part of the answer to the question. I would look at the Leapfrog CPOE consortium and become involved in that.

The Roadmap for National Action on CDS, that I mentioned earlier is a step in trying to bring together key stakeholders such as the ones that you described to do just the sort of thing that you're talking about, getting the vendors together, the implementers, the payers, providers all sorts of folks like that. And there have been a lot of positive results that have come out of that in terms of early steps in moving the ball forward and with the change in administration, I think there will be more opportunities going forward. I would turn this over to Dr. Friedman to comment on the notion of bringing together the different stakeholders and things like the AHIC CDS Workgroup have a lot of conversations about.

Sure and I actually wanted to mention one thing specifically. The Certification Commission for Health Information Technology, CCHIT, has taken interest in a new approach in what they do. Certification as you all know is directed at the floor and making sure that minimum standards are met. The CCHIT is considering an Advanced Technology Program and instead of focusing on the floor it will focus on the ceiling and incentivize vendors so that they earn a merit badge in specific technological areas. There is no indication that one of these will be CDS but it will be a candidate for such a program out of the CCHIT. And among the things that I have been thinking about, that would address this question of improving CDS offered by vendor products; it would need this kind of an incentive. So, I am personally very excited about this and I think it can have a very good impact in the very near future.

Thank you Dr. Friedman and thank you for that informative discussion, we will move on to Randall Cebul and Peter Greco.

Thanks. This is Randall Cebul and I will go first.

I think we will pursue this metaphor of powerball and smallball. We were asked to talk a little bit about a powerball trial that had a lot of smallball evaluations that were part of it. So I will try to spice my presentation with some allusion to those concepts. As a summary, success, we're talking about measures of success. And I think our success with CDS can be measured in several ways. Whether a component was ever used, a concept of adoption, which I will describe in detail, user satisfaction, improving care processes, improving intermediate outcomes and real outcomes of patients. So I think there are a number of attributes of success. And obviously the improved real outcomes of patients with an effect size that can be successfully measured is this sort of powerball approach we've been talking about. The second targets of CDS may be patients, providers or other actors in the system. Patient level success should consider what you're trying to accomplish. I will show you later an example where the workflow dictated that the best approach for CDS was actually in the earlier stages, such as upon registration at an outpatient clinic.

Further, CDS can influence provider behavior and processes, and, as I will show the powerball results of our trial, they may be less likely to influence patient outcomes. And finally, I will allude to this but Peter Greco will elaborate on it and that is that alert fatigue is a treatable condition and filtering can improve specificity and increase the likelihood that people will respond to alerts.

The trial that we did is what we call DIGIT, which is Diabetes Improvement Group Intervention Trial, it was a cluster trial, a large trial, to improve diabetes care and outcomes. We will describe, briefly, some success in CDS measures and some of the design features of the trial that were intended to increase success in some of our powerball results and system related CDS and the results of this workflow sensitive intervention.

I think this definition has been provided before. The only thing that I would say here is that the right person might not always be the physician.

These are the patients; this is a map of greater Cleveland. The river that burned is the white space in the middle and Lake Erie is in the north. And these are about 14 thousand patients in the greater Cleveland area that are cared for by 200 physicians in about 24 different practices. This was a cluster randomized trial in two systems. And I will focus on the one system, the Metro Health System on the left that uses an electronic medical records system. As Dr. Friedman pointed out in his powerball analysis, we tried to freeze the decision support that was available in the Epic system only and in the DM2 arm, which is Disease Management for Diabetes. We basically introduced the decision support that I will describe very briefly. So there were 10 practices and about 65 doctors and 6 thousand diabetic patients in this trial.

The CDS that I will allude to here includes real time alerts and linked order sets. The alerts were filtered and Dr. Greco will describe them later. There were patients and physicians learning what we call practice panel tools which included patient registries for every doctor and weekly updated performance feedback on the individual physician's practice with regard to the metrics that we employed and their performance in comparison to their peers in the group practice or throughout the system.

The success measures that we included in this trial included the concept of adoption. Are the alerts adopted? Or were they ever used? Of all the alerts that were triggered in the patient encounter in the office, was the appropriate action taken for that opportunity? Here we have a denominator, which is the number of alerts that were triggered and the numerator is, did the physician take the action that was implied by the opportunity? The second, at the end of the trial, this is another sort of smallball question. I think Dr. Friedman said, what's the buzz? The question is, at the end of the AHRQ funded trial, we

basically were able to keep all or most of these registries and other kinds of decision supports up and running with a relatively small incremental amount of effort. But we did not know whether, in fact, if this was something that the physician would find a good idea or enable them to get rid of a burden. We prospectively asked those questions on whether or not the CDS, both individual components as well as an aggregate, should they be kept after the trial is completed? Third we looked at the differences in care processes in diabetes, did they receive timely hemoglobin tests, LDL cholesterol, urine microalbumin, pneumococcal vaccines, and receipt of ACE inhibitors or angiotensin receptor blocker drugs and then differences in good outcomes, and I would say these are more patient center metrics. And are none the less the intermediate outcomes as opposed to amputations and blindness and so forth.

So those were the metrics that we used. The smallball measures related to adoption rates and to whether the providers like the various components of the clinical decision support as displayed here on this slide. And overall in the left-hand figure, the overall adoption rate was 28%. That is out of all of the multiple thousands of opportunities for responding to the particular alerts, 28% were taken. This ranged from 48% at the high end, reminders to do hemoglobin A1C tests, at the low end the alerts that enabled the provider to refer a patient using an automated order set to a nurse case manager.

At the end again, we asked the practice sites that had the clinical decision support whether they wanted to keep the alerts and order sets and whether they wanted to keep the panel tools and the nurse case management supports. And the answers here were actually quite favorable. And even to the point that they wanted the nurse case management alerts kept even though we couldn't afford to keep the nurse case managers.

In terms of the powerball analysis, the red bars here are the intervention group or the CDS group and the blue bars are the usual care frozen arms of the trial. The answer was we did not really change outcomes. The odds ratio of meeting all of the outcomes was no greater in the CDS group then in the control group and there was a 50% higher odds and only a 6% actual percentage increase in the processes in CDS group as compared to the control group. So I would say there is a significant secular trend in the favorable direction both for outcomes and processes and the processes and the powerball analysis allows us, at least, to differentiate those things that might be more responsive to CDS as opposed to not. So I think we might have, with sort of a secular trend analysis and no comparison group, been misled into thinking that we had made improvements when, in fact, we perhaps had not.

These are the specific components of the process standards that were changed.

The first summary that I would make is that there are some reasonable soft or smallball measures. The adoption is not something that we have seen in the literature but we think this is probably a clue, something that is accessible to us, as to the meaningfulness of the alerts and the reminders in the context of the visit. So I think this is a reasonable metric. Provider satisfaction with this, keep it or not, is a very nice and simple smallball measure that caused us, in fact, to not only keep it with the intervention group practices but to also disseminate it out to the 12 practices and our system.

I think the useful alerts can enhance the care practices by doctors and comparison and control groups are useful to show whether it is greater than what would have been seen by secular trends. And the alert to primary care providers might not include outcomes and the example I give is that most primary care doctors are likely to be aware of their patients' poor hemoglobin A1C value and remind them that they should order a test or intensify treatment but may not cure the problem.

Along the way, we thought that the pneumococcal vaccination rates were not good in our diabetics and other patients who have the conventional indications for it. And what nursing actually did is work with Dr. Greco and others to identify scheduled patients who meet criteria for the vaccine and have not received it and then the patients were identified by age and the diagnosed conditions and that list was provided to receptionists and the nurses and doctors basically enabled standing orders to be written for a nurse to offer and administer the pneumococcal vaccine before the visit.

On this next slide we are now in the process of a region wide quality improvement and performance measurement initiative. In the top, practices with regard to pneumococcal vaccine rates among diabetics in greater Cleveland are part of the system that has adopted a workflow sensitive alert system for the receptionists and the nurses and it has been a remarkable success. And absent data on the sites we would not have been able to say how great this success had been.

This is Peter Greco taking over for the remaining six minutes or so.

I will talk about alert fatigue as a treatable condition and expand on some of the decision support that Dr. Cebul described.

This slide shows an example of two of the alerts that had fairly low adoption rates. The first alert, consider an ACE or ARB for a patient with microalbumin of 30 or higher, has an adoption of I think 14%. The second was to consider Statin for LDL of 100 or higher and had an adoption rate of 11% or so-- by the powerball analysis these were not terribly effective. But one other measure of success for these is, did we accomplish what we intended to? Did we avoid false positive alerts? And I think we did because these would not display unless several conditions had been met. These would only apply to diabetic patients. They would only be displayed for patients who had leaking protein in the urine or in the case of the second alert, had an LDL cholesterol above 100. These alerts would only display if patients were not already on the treatment and further, that they were not documented to be allergic to those medications. And finally, that there were no obvious contra indications for these interventions. The last thing we want to do is make a recommendation that would harm the patients.

If we delve more deeply and show these alerts in the context of the visit in the EMR, towards the bottom, there is a link that they could click to access the alerts. And how often did they go in and do that in response to seeing this alert? This shows the example of a busy screen that we have made busier with some overlays. But at the top it shows that we are giving the user some selections or options for medications to be prescribed to this patient. Some tests that could be ordered and so on. So we've found that those alerts did not alter the outcome in the powerball sense but the users thought that we should keep them and they thought that we were doing something useful. So they remain in our system to this day.

I will switch gears and talk about how we looked at a different set of alerts. This is not related to our DIGIT trial. This has to do with a related issue of alert specificity and improving the response to alerts. And this is probably a classic example of what you call a smallball analysis. Because when looking at drug interactions, the ultimate goal is to prevent drug interactions and that is not what we studied. We studied, did we filter out the alerts that we wanted to filter out and did they pay attention to those alerts? And the background to this is that we all knew that anyone who used our system had a lot of nuisance alerts going off and we wanted to do something about that and our particular third-party vendor categorizes drug interaction alerts by severity and likely documentation level. In other words, how likely is

this to be a true problem? So there are 15 possible combinations of alert severity and we filtered out the ones that we thought were least important, i.e. the least severe and least well documented. I'll skip over the methods, you can read them at your leisure, but basically we looked at what alerts were going off most often and we had generalists and specialists review those and figure out which ones to suppress based on the ones that we thought were nuisance alerts in the system. This is one of our smallball results which is simply, did we succeed in filtering out the alerts that we intended to filter out. And this graph over time shows a dramatic decline in the rate of drug interaction alerts displayed to end users in their use of the system and that that decline was almost entirely made up of fewer minor alerts, which is the green line, which is almost zero and few were moderate alerts, which dropped dramatically. And really no change in major alerts which is the red line at the bottom. And the net effect of that was an increase in the average importance of each alert that was presented. So we consider that outcome a success.

The other outcome we looked at is how did the users respond to those alerts? Did they actually proceed with prescribing the interacting drug? And prior to filtering, the cancellation rate, in which the user would cancel the order was less than 1%. After filtering, you notice we are displaying about 75% fewer alerts. And the response rate was more than three times as high although still a very low rate overall. But the point being that the users paid more attention when we showed them fewer alerts but of higher severity.

And if we look at the major severity alerts specifically you notice that these are higher at baseline than it was for all alerts, which we think means that users were paying some attention, but that increased as well.

Our conclusions were that with fewer alerts displayed, a much greater proportion of the alerts were attended to. And we believe that this represents reversal to what is commonly referred to as alert fatigue and have some promise for improving the usefulness of alerts in these complex systems.

With that I will conclude our discussion. Thank you.

Thank you Drs. Cebul and Greco, one point before I articulate a question. A request was put forward with respect to a citation of one of the info button articles that Dr. Osheroff referred to and we will add such references into an augmented list of references that will be available for this webinar.

We received a curious question that is generating discussion.

The commenter says, I am curious about what the panelist thinks about how far or close we are from developing standards for CDS evaluation that would enable both easier development of evaluations by others as well as easier comparisons, meta analysis, etc. It seems like there are specific types of questions that keep coming up and I wonder if any groups are trying to standardize evaluation metrics and tools for CDS.

We will turn to Dr. Cebul and Dr. Greco for their first response, if they care to.

Sure, I guess when we designed our trial, we did a thorough search of the literature and spoke with colleagues elsewhere. And I think that we've found that a number of the metrics that we were looking for were not as sophisticated as we felt we could be. This concept of adoption, for example, is something that is a clue that can address issues of alert fatigue. And even as Dr. Greco said, 14% is the adoption rate for the ACE/ARB alerts. That actually is probably a relatively high adoption rate if you can think that

somebody perhaps had a cough and some said the medication might not be in the patient's record. I guess I would say that there's room to improve and standardize some of our metrics.

Yeah, the thing that I would add to that is that some of these metrics are going to depend on the system in which you are implementing your alert. For instance in our case, there are certain actions that the electronic medical record captures that you can report on and certain measures that you cannot report on. We do not know every action that the user takes in every encounter. But, for example, we can know that they canceled a medication order after being displayed an alert; we can use that as a measure of adoption or lack of adoption.

This is Dr. Friedman. If I could jump in here. I've thought for a long time and continue to believe that most evaluations are local and the generalizability of tools and methods from one study to another is limited for some of the reasons that we just heard. On the other hand, I do believe in the spirit of the question that was asked, that we could take some significant steps to develop reusable templates for evaluation methods and metrics related to CDS that would really keep everybody doing studies, particularly smallball studies from having to start all over again. There is some experience in doing this with an evaluation tool kit that AHRQ developed. Although this wasn't, if my memory serves me, targeted on CDS, I wonder if any of the AHRQ people participating could comment on the experience with the Health IT Evaluation Tool Kit.

Thanks Dr. Friedman, I'm just pulling up the toolkit. You're right that it was not specific to CDS, it was developed for a broader health IT evaluation. There is one for health IT and another for health information exchange. Another project that is ongoing though is for the health IT toolkit; NRC staff are developing metric briefing sheets for metrics with a lot of evidence in the literature. We're developing a three-page guide on how to use and implement those.

I would add two quick things to the excellent comments made by the other panelists about this question. First of all in terms of the importance of having standards for CDS evaluation, I can tell you that the 80 or so people who contributed to the Roadmap for National Action on CDS identified this as one of the pillars of things that needed to happen going forward. That document was released two years ago and it made the exact same point that is being made here, that this is something that is really critical and there is a lot of work to be done there. That has not been picked up as far as I know in terms of specific efforts of moving forward but I would comment in that chapter seven that we talked about in the new guide, we laid out specific recommendations about what to evaluate and how to evaluate. My hope is that kind of information will serve as a framework, not for sort of formal standards, in an HL7 or a CCHIT sort of way, but at least in a more generalized way of thinking about how we are evaluating and what we are evaluating so we can have a lot more of this kind of information available for sharing through things like the Wiki and other publications and stuff like that.

We have a few other questions, one is asking for clarifications with regards to the smallball evaluation, is it the same or different than quality improvement measurement approach?

Dr. Friedman?

In terms of underlying philosophy they are very similar. Both are extremely pragmatic, results oriented and decision directed. I could go into an exegesis here on different approaches to evaluation and the philosophy that underlie them, it suffices to say that what I call smallball and the quality improvement approaches to evaluation derive from the same set of assumptions about what is important in evaluation

and what should get emphasis. And I think the person who asked the question appropriately picked up on that.

Dr. Friedman, would you care to elaborate on the struggle that researchers encounter with respect to getting published and competition with other powerball applications, do you have any guidance?

Yes, and here I am drawing on partially my experience as an Associate Editor of JAMIA over several years. And my general sense of the way the informatics field is going. I really do feel that there is an evolving and growing sense of appropriateness of studies done in context in a larger project to which they contribute to which they enhance. One example is that JAMIA has several publications types that really would be compatible with project descriptions and results reported as results of smallball studies. And I also think, and here I would emphasize the fact that smallball studies are careful, rigorous studies. Just because they are small, does not mean they are sloppy or anecdotal. I think looking at the field at large, kind of from the other direction, I think there are several meeting the rigor of which would benefit from the kinds of smallball, carefully done small studies that I am talking about and that particularly the last presentation illustrated in lieu of much more anecdotal and less careful studies and data that are used as the basis of these presentations. So think about smallball as kind of in the middle between powerball studies and anecdotal evaluations. And there are a lot of domains and situations where smallball would be a step in the more rigorous direction that is needed in my opinion.

Thank you, Dr. Friedman.

We are almost out of time; I will go round to the panelists to see if they care to add any information, Dr. Osheroff?

Yes there was one question that I would like to respond to and I think it's a nice wrap up to the comments that I was trying to make. It’s about using business process management software for implementing the five rights. I think what I would say there is what we're trying to do with CDS is sort of the Swiss cheese model for patient safety, I think the idea is to lay a lot of different approaches out there so that the bullet will get stopped somewhere in the Swiss cheese and not make it all the way through the holes and I would encourage people to think about their CDS evaluations and efforts in that way as a series of things to make sure we're getting to where it is we’re trying to go.

Thank you.

Dr. Cebul?

Is there anything else that you care to add as conclude our discussion?

Dr. Greco?

Okay.

I thank you all for your time. In about a week or so the PowerPoint presentation and a transcript will be available as well as an annotated bibliography for your reference. And we thank the panelists very much.

They hit it out of the park, smallball or not.

Thank you everybody.

Thank you.