FOOD AND DRUG ADMINISTRATION

SENTINEL NETWORK PUBLIC MEETING

March 7, 2007

University System of Maryland
Shady Grove Center
8630 Gudelsky Drive
Rockville, Maryland

Proceedings by:
CASET Associates, Ltd.
10201 Lee Highway
Fairfax, Virginia 22030
(703)352-0091

Table of Contents

Table of Contents (cont.)

P R O C E E D I N G S

Agenda Item: Welcome and Opening Remarks

DR. SHUREN: I'm Jeff Shuren. I'm the assistant commissioner for policy at the FDA. I would like to invite all of you to the Sentinel Network public meeting to promote medical products safety.

I am delighted that so many people could attend, particularly in light of the weather. I think folks know that when there is even a threat of snow in Washington, whether it be to you or a family member that doesn't even live here, you can shut down anything. In fact, that is happening. There are schools closing all over the place. They were closing yesterday, even though there was no precipitation. So thank you for coming.

On behalf of all of us in the federal government, we want to thank you for participating and we look forward to a fairly dynamic meeting.

I would like to begin today's meeting by introducing the Commissioner of Food and Drugs, Dr. Andrew von Eschenbach, and the deputy commissioner and chief medical officer of FDA, Dr. Janet Woodcock. Dr. von Eschenbach is going to be giving opening remarks, followed by Dr. Woodcock, who will describe the vision of the Sentinel Network.

After those presentations, I will review some

housekeeping issues -- most importantly, where you can get lunch. Then I will introduce the federal government panelists and the invited speakers, and we will get started.

Again, welcome.

DR. VON ESCHENBACH: Thank you, Jeff. Good morning, ladies and gentlemen.

I want to begin by, first of all, echoing Jeff's remarks. The very fact that I see so many of you here, committed and really dedicated to the effort and the purpose that we are here today to discuss -- namely, to create an opportunity to bring us together in a network that will enable us to serve patients and the American public in ways that, perhaps, 5, 10, 15 years ago, we couldn't even imagine -- is a tremendous testimony to you. It's a testimony to your dedication and your commitment. For that, we are extremely grateful.

It is also gratifying to me because I have a very strong belief in how critically important your effort and your participation is to what I believe is perhaps the most profound transformation to ever occur in the history of medicine. We, as we begin this 21st century, have an opportunity to radically change not just our health-care system, but our whole concepts of health care and of health itself. It is because of the tremendous progress that has

been made in science and technology that we are able to now begin to discover the very fundamental mechanisms that are responsible for disease, and responsible for health. That is leading to the creation of interventions, solutions that will enable us to do things that previously seemed impossible.

But one of the most important parts of that discovery and that development is, in fact -- obviously important, because until all of that progress actually touches someone's life, it really has not found its meaning or its purpose.

But even more importantly, as we look at the implications of what is occurring in science and technology, as we look at the implications of the fact that we are moving from a macroscopic and microscopic perspective to a molecular perspective, what we are beginning to envision and see is that the delivery end of the continuum is not just an opportunity to provide and deliver these solutions, but in the very delivery of those solutions there is the development of new data, new information. If we have the ability to extract that data, acquire it, assemble it, analyze it, we in fact have an opportunity to rapidly and radically transform the evolution of knowledge -- knowledge of what is happening with regard to disease and health in the human condition,

not in the laboratory.

One of the most important, critical factors associated with what is happening when we apply these solutions to patients whose health depends upon them is, are we providing to them solutions that are, in fact, as effective as they were intended to be, and are they as safe as we believed them to be when we developed them and discovered them, in a very confined laboratory or clinical-trial setting?

Ten or 15 years ago, that was a question that perhaps we could not answer, because we didn't have the tools; we didn't have the technologies. Today those tools and those technologies are available to us. They still need to be further developed. But we clearly can see enough progress in that regard, and enough progress that has been made in other industries in which the ability to acquire and assemble data has now become routine -- whether it is what is happening with your credit cards and your bank accounts or your cell phones -- now we have the opportunity to do that in health care. We have the opportunity to bring us together to create a network and a community so that we can begin to acquire the data to allow us to have the information that we can convert into knowledge.

What better way to begin than to do it in the

context of understanding safety so that we can protect and enhance the lives of those we are serving? That is your work today, to begin to come together as a community to discuss the Sentinel Network, a virtual, electronic, integrated medical-products safety system. It can be one of the most profound contributions to the new era in health care in this nation, and ultimately in the world.

I am very grateful for our partners, partners who have assembled from other parts of the federal government: the Agency for Health Care Research and Quality, the Centers for Disease Control and Prevention, the Centers for Medicare and Medicaid Services, the Department of Defense, the Office of the National Coordinator for Health Technology, the Veterans Administration, the National Institutes of Health. The important message of not just thanking them, but listing them is to demonstrate to you that we are committed. We are committed, as federal agencies, to work collaboratively and cooperatively together to bring this new reality to the fore.

We want to collaboratively and cooperatively work with you. We want to listen and learn from you and understand both the challenges and, more importantly, the opportunities that we together, as a community, can begin to seize and implement.

It is an extremely important challenge. It is an

opportunity that will change the lives of millions of people and save lives and enhance the safety of our medical care system, improve the quality of the products that we are providing to those patients, and, more importantly, continue to assure the good health and well-being of this nation and of the world.

You have important work before you. It's a privilege for me to come and welcome you and thank you for that effort. I wish you well in the venture, on behalf of all those lives that you will be affecting.

Thank you.

Agenda Item: Vision of the Sentinel Network

DR. WOODCOCK: Good morning. I'm Janet Woodcock.

I am going to talk a little bit about where we see the future going and why we have convened this meeting at this time.

The background on this is something that I think everyone in this room -- although certainly, by all means, not everyone in the nation -- these are truths that we hold to be self-evident, so to speak. Adverse events due to use or misuse of medical products are very common. Of course, that part is always in the paper. Adverse events, as you all know, may result from the inherent properties of a medical product -- in other words, they are what we call the inherent properties -- or a problem with the

manufacturing of the product -- we see that more commonly with some sorts of products than with others -- or, probably most commonly, errors or perhaps just slight misuse in prescribing, selection, or use. That is most common, probably, with the pharmaceutical end of things.

We know that, adding all this up, there is a tremendous toll of harm inflicted on the population, based on the use of medical products. But it isn't that anyone is trying to inflict this harm. We are all trying to do the best we can, each part of the system. But right now we think health-care practitioners and patients actually do not have all the up-to-date and accurate information about the benefits and risks of the medical products, especially at the point of care, that is needed to use them safely and effectively. We are not using the information we have now and applying it to the proper care of patients, and we are not collecting the information as rapidly and efficiently as we all would like to. I think there is agreement on these points.

The benefit/risk profile of medical products evolves over time. When we first start using a product, whether it's a device or a vaccine or a drug, we know a certain amount about it. But over the next decade, or even the next 100 years, we continue to learn more about the products.

The premarket clinical trials that are done, of course, to get products onto the market can't identify all the potential risks, as we know, or even the future changes in use patterns that are going to occur once that product gets out there. Therefore, full knowledge of benefit/risk doesn't exist at the time of product approval, and often not at the time of treatment decision making. As I said, often even the full weight of information that we actually have is not brought to bear on those treatment decisions, because that information isn't in the hands of the prescriber in a usable form at the time of the prescribing or use.

As I said, when this information for safe use exists, it's not always readily available at the time of treatment decision making. Therefore, timely and effective postmarket surveillance and risk communication -- in other words, communication of the evolving information -- as rapidly as possible are critical to reduce the knowledge gap and foster better-informed treatment decisions and actually keep people safe.

The effectiveness, however, of the federal government postmarket surveillance and risk-communication efforts, along with the industry efforts, has been constrained due to the following limitations:

- Number one, the quality of the data that we

have to deal with.

- The quantity of data. Often we only see a very small fraction of the harm that is occurring that is actually reported in some manner.
- The timeliness of acquiring this data from the health-care system and being able to analyze it.
- Our capacity as a system -- the federal government, the industry, academia, and the health-care system -- to rapidly conduct postmarket studies when needed to gain the information, to confirm signals that have been arising.
- The risk-communication tools that are used. When I started at the Food and Drug Administration, the risk-communication tool was a letter to the doctor, the "Dear Doctor" letter. We still use that, to some extent. It is not an effective communication tool. That is not to say that anything is being done wrong. It is saying that we can do better than that now.
- The resources available to all these sectors to actually conduct all this work and to improve the situation.
- Our current efforts are frequently crisis-driven rather than prevention-driven. We identify signals and we must go out rapidly and do studies or try to figure out what is going on.

- Our analyses currently focus on overall populations rather than differentiating subpopulations. In other words, we, in many cases, do not have the capacity to connect the basic science with what is happening out in the clinic, to say, is this adverse event due to differences in drug metabolism, is this adverse event due to some particular drug interaction, et cetera? We aren't bringing all the available science that exists right now to our analysis.

- Most importantly, I think, there really is no coordinated systems approach this. We are operating off of systems that were put in place many years ago, before computers really were in wide use. We have computerized our data collection in some cases or our databases in some cases, but we haven't really taken a very hard look at what the new world that we are moving into could do for this situation with adverse events and with the harm that is being inflicted on the population.

From the FDA's standpoint, and I think from the federal standpoint -- our federal partners who are here -- we see that the private sector is moving ahead and taking steps that really can facilitate surveillance activities and help us learn much more rapidly from the health-care system. They are developing new information-technology tools, through the e-health record and so forth. They are

exploring informatics methods and analyses. A number of the health-care systems are performing their own analyses to try to improve quality within their own organizations. This is creating the capacity to conduct within the health-care system, in a new way, postmarket safety assessments, in a way that has not really been done before.

Therefore, we feel we need to think and discuss how to link the private and public-sector efforts together in the future, or have a plan to address the limitations that I have gone over, through better integration of the nation's postmarket medical product safety activities (the private activities that are growing, the public activities that are happening in many agencies, the industrial activities), to create what might be called a Sentinel Network -- in other words, a virtual, integrated, electronic nationwide medical product safety network -- so that we all are able to use the efforts of each other to the common good of protecting patients and improving our knowledge about medical products.

We would like to aim toward a network that would foster seamless and timely electronic flow of information about medical product safety from what is out there in the health-care arena, as well as from the surveillance reporting systems, through various risk-identification and risk-analysis procedures. Some of these are being done, as

I said, in health care. Some of them are being done by the Food and Drug Administration. Analyses are performed by the sponsors of these products, and so forth. We need to make sure that information is all put together and is transmitted to practitioners in a timely manner at the point of care so that they can practice with the best possible information.

We are not advocating assembling some large, grand database or anything like that. What we are talking about is to have a public-private collaboration or partnership to build on existing efforts and to connect them, not to create new databases or structures. We would hope that the network could use national and international standards that have been adopted by AHIC, but also that have been developed through national standards organizations.

Components that we need to look at today -- and I think I need to stress that today is going to be our first and more or less preliminary exploration of this:

We need to look at data collection, who is collecting data where and how, and where is it residing. Electronic health records are an obvious opportunity. They capture adverse events of different kinds in the context of medical product use and in the context of clinical practice. So EHRs or other types of integrated databases

that might be used with the EHRs are sources of data that we need to talk about. There may be other sources of data out there. We know that there are registries, large registries, that are conducted in some academic settings and so forth, and we need to talk about those.

Risk identification and analysis: The folks here at FDA tell me repeatedly that one of the things we need most right now, as we begin to get all these data sources, is research into how to analyze them correctly and figure out the wheat from the chaff.

We need integrated research networks, and we need to find out which ones are going on where and what they are doing.

People are discovering data-mining tools. Obviously, data mining is very advanced in some other sectors, such as defense and so forth. The question is, what research and effort do we need to do to apply that type of data mining to this problem?

We need eventually to reach agreement on some methodologies. Obviously, at some level, these analyses and decisions reach a level of controversy, and to have methodologic agreement will be extremely helpful for all of us, I think, to reach some kind of consensus on proper methodologies and how you can draw inferences from these types of data.

How to integrate biology into all of this is something I am very interested in, not just empirically looking at all this, but whether we can actually have ways whereby signals and hypotheses that are generated out of this system are actually translated into the biological meaning. Is there an underlying mechanistic explanation? Can we do research to find that out? That sort of backwards translational research is going to be extremely important. In fact, I think if the promise of all this comes to fruition, it will be by connecting this back to the biology, so that we actually learn knowledge, not just develop information. We need to understand why these things are happening so that we can prevent them.

Finally, equally important -- and I don't know how much of this sector is represented here today -- are the risk-communication aspects of this. Our system right now for communicating to physicians, other health-care practitioners, and patients is extremely fragmented and isn't working very well, to be very blunt about it. We need to not only leverage the medical community's expertise, we probably need to reach into expertise of those who are much better at communicating. To communicate to clinicians, this new risk information will have to be integrated into the workflow of clinical practice. There is no doubt about that. We cannot simply bombard

physicians, pharmacists, nurses with pieces of paper giving them information. We need to integrate that information in a way that is usable to them during practice.

We know that new decision support systems are being developed out there. That is another piece that we all need to think about in a Sentinel Network: How do we effectively transmit this information or make it available to people?

What are the objectives of today's meeting?

- We should be able to, hopefully, by discussing across a wide number of parties, evaluate the current needs in postmarket medical product adverse-event data collection and risk identification and analysis -- not only the needs; we should be able to find out, I think, from a wide variety of stakeholders what people are actually doing out there, what they would like to do, what the gaps are between what could be done, what people would like to do, and what is actually going on right now.
- Then identify the obstacles to and what could facilitate or incentivize developing the components that we need to close this gap and connect these various efforts up into a Sentinel Network.
- As part of that, we can identify opportunities for public-private collaborations, for assembling the data and for doing the research piece and the analysis piece,

the risk identification and analysis components, of the Sentinel Network.

Obviously, some parts of this belong in academia. They are very research-oriented. Some parts are analytic and perhaps belong within the federal government and the industrial sector. For other parts, we need to figure out who would do them.

After this meeting, the public docket will remain open until April 5. We will review the oral and written comments that we receive on this. Then we will attempt to develop from this a roadmap for assembling a Sentinel Network. This might include, of course, additional meetings, formation of public-private partnerships, other steps that we might need to take. Until we hear from you at this meeting, we are not going to know exactly where we need to go next.

Thank you very much. I will turn it back over to Jeff.

DR. SHUREN: Thank you, Dr. Woodcock.

Agenda Item: Meeting Format

Now I would like to go over some housekeeping details for the meeting.

We have our federal government agencies represented here at the head table. In a minute, I will introduce the folks who are up there.

We have also set aside seats in the front here for our invited speakers. Each of those invited speakers is going to give a 10-minute presentation and we are going to have five minutes of Q&A with the federal government representatives.

Those individuals who have registered to speak will be giving their presentations in the afternoon.

Let me just mention that some folks are having difficulty getting here because of the weather. Some people are flying in. Some of the folks who are registered to speak in the afternoon -- if we move a little bit more quickly, I may ask them to present this morning rather than in the afternoon. We will just try to be a little bit flexible with the schedule.

For those who are registered to speak, we have also provided seats for them off to the side. When folks do come up to talk, we will try to move it along quickly. I will ask you to stay within the timeframe, just as a courtesy to others. Those who have registered to speak will have five minutes, with three minutes of Q&A. I do have a person sitting up front here who will let you know when you have two minutes left to speak. We will signal this and we will let you know when you are out of time. No one is going to throw rotten tomatoes, but I just ask that you try to stay on schedule.

Tomorrow we are going to hold a moderated dialogue between the federal government and invited speaker panelists to discuss concrete steps that we, together -- the public and the private sectors -- should take to assemble the Sentinel Network. We have also included times throughout the day for the rest of you to participate in that discussion. We are going to set up a rectangular table in the middle here and have federal speakers and invited speakers sit around that. Everyone else is going to be around. We are going to encourage you, at selected times, to get up, give us your input. If you have other thoughts and ideas, we want to hear about it. This is really meant to be very, very interactive. We do want to hear your thoughts.

Again, I remind you that the public docket is open until April 5. If you do have any written comments, please send them to us.

We hope everyone has registered. If you have not, please do so. It just helps us keep a record of who has come to the meeting.

We are going to be breaking for lunch at noon.

[Administrative announcements]

Lastly, I just want to note that a transcript for the meeting and all presentations will be posted on the docket on the FDA Web site. I ask you to just please be

patient; it's going to take a little time to get the transcript done. But we will get it out there.

Now I would like to introduce the federal government panelists. Not everyone is here. Some folks are having difficulty getting in. But let me just at least mention them, because I do want to make sure you are aware of the organizations that are involved.

From AHRQ, we have Drs. Jean Slutsky and Anne Trontell.

From CDC, Dr. Dan Budnitz.

From CMS, Dr. Jeff Kelman will hopefully join us.

From the Office of the National Coordinator for Health Information Technology, Ms. Kelly Cronin.

From the Department of Defense, Rear Admiral Tom McGinnis and Lieutenant Colonel Mike Datena.

From FDA, we have Dr. Gerald Dal Pan, Dr. Tom Gross, Dr. Miles Braun.

From the NIH, Dr. Clement McDonald.

From the VA, Drs. Fran Cunningham and Mike Valentino.

For invited speakers, again I will introduce some folks who are not here at the moment. A number of them will be here tomorrow.

Dr. Jeffrey Hill from the American Medical Group Association and Anceta.

Dr. Ken Mandl from the Center for Biomedical Innovation and Harvard Children's Hospital.

Dr. Marc Overhage from the Indiana Health Information Exchange and the Regenstrief Institute.

Dr. Rich Platt on behalf of the HMO Research Network, Harvard Pilgrim.

Dr. Hugh Tilson is going to come to speak on behalf of Dr. Rob Califf, who is at Duke. Rob will be here tomorrow, but Hugh is going to give his presentation and talk a little bit about the CERTs.

Dr. Chris Chute from the Mayo Clinic.

Dr. Barb Rudolph from the Leapfrog Group.

Dr. Michael Caldwell from Marshfield Clinic.

Ms. Liz Paxton from The Permanente Federation.

Dr. Fred Resnic from the Massachusetts E-Health Collaborative.

I also want to express regrets. Two folks were not able to make it at the last minute, but they are part of discussions here. Dr. Garret Fitzgerald from the University of Pennsylvania and Dr. Wilson Pace on behalf of the Federation of Practice-Based Research Networks.

With that, again let me welcome you. Let me introduce our first speaker, Dr. Jeffrey Hill.

Agenda Item: Presentation by Jeffrey Hill

DR. HILL: Thank you, Jeff. Good morning,

everyone. Thanks for coming. It's a pleasure to be here.

This morning I would like to introduce you to the American Medical Group Association and its Anceta Collaborative Data Warehouse. I want to give you a quick overview, which will set the stage for some further comments and questions, I hope.

Anceta is the health informatics subsidiary of the American Medical Group Association, referred to as AMGA. It supports AMGA's strong tradition in collaborative quality improvement.

Before I start talking about Anceta, I probably should say a few words about the American Medical Group Association itself. The AMGA represents the leading multi-specialty medical groups around the nation, nearly 300 medical groups, representing about 80,000-some physicians in 42 states, representing the care of about 50 million active patients across the country. This is done as a team sport. These are coordinated care-delivery systems, large and small, and medical groups involving primary care, as well as specialty care. Many have both inpatient and ambulatory care.

You can see here, just by looking at the names of the current board of directors, the scope of the types of members that we have, from the large integrated health-care systems, such as the Cleveland Clinic, Kaiser, Geisinger,

to some smaller ones, represented by Everett Clinic in Washington, Mount Kisco in New York. So it's quite a broad spectrum of members around the country, also representing academic medical groups in a number of medical centers around the country.

This is just a quick map to show you the spread. They represent both rural and urban care, pretty much around the country -- not all parts of the country, but certainly representative. There are a couple of groups even in Hawaii, which is not shown here.

This was formed a few years ago as a result of AMGA's number of one-off studies looking at outcomes research and quality studies, sponsored or not, where groups would come together and share their data. Every time they did that, anonymous to each other, just by looking at the data among themselves and others, their level of quality would improve. At one point, though, the board of AMGA said, "Why are we doing these one-off studies all the time? Why don't we form a national data warehouse so we can slice and dice this any way we might want," because research questions, as you know, and the results that they generate always generate additional questions, and you may or may not have the data to go to the next level of query, unless you have access to the full set of data.

So our objective, on a mandate from the AMGA board, was to put together a national collaborative data warehouse, with the objectives shown here:

- Mainly to give access to each group that is participating to their own data, be it administrative, ancillary, clinical data, in a standardized, analyzable format for managing quality cost and utilization. The idea is that everyone is looking at their data in the same standard format with the same kinds of tools, so they can, in the tradition of AMGA, share the knowledge around the data and share best practices in improving care to their patients.
- The real emphasis of this, though, is to have comparative data among themselves. This is not about connecting or collecting data within a group, but it is more about collecting, aggregating, and analyzing the data among multiple groups around the country who are looking to get comparative data and benchmarks in areas such as the ones listed here, whether it is practice management, looking at the performance of the physicians or products, be they drugs, medical devices, or biologics, and certainly to continue to look at health outcome studies, the economics of care, and to continue in the tradition of improving the quality of care.
- There also are a number of treatment and

outcome studies that can support evidence-based medicine that we will hope will get back into their workflow, whether it is through their decision-support systems and their EMRs or not -- of course, to improve the process of delivery of care in a coordinated environment.

- Ultimately, to have a large, rich database for health services research among them.

There are also a number of standard tools. I say "standard." Everybody uses the same kinds of tools, whether it's for clinical trials, looking at protocol modeling, or looking for cohorts of patients, whether it's a drug or device trial, or maybe just a quality-improvement or disease-management program, and also to look at results-based performance measures and how we can use these for evidence-based guidelines to improve accountability.

Up front, we all understood that there was a need to underwrite this exercise. So we are providing, through third parties, a new source of health-care information to all segments of the health-care and life-science industries. We can talk a little bit more about that later.

As I said earlier, this is really enhanced by the types of relationships among these multi-specialty group members. We are able to collect and aggregate and analyze comprehensive patient-care data because of the very nature

of the medical groups that are contributing this. They are a coordinated care-delivery system involving in most cases not only primary care, but most specialties, including in many cases inpatient facilities, as well as outpatient clinics and ambulatory care. There are both in-house and external lab facilities or access to the lab data, and similarly in pharmacies. These are large enough to either have inpatient pharmacies -- and many of them own their own outpatient pharmacies -- or have relationships with pharmacies, whether it's through their EMR or PBM relationships, to get access to the pharmacy data -- and, or course, radiology and the like.

The other thing that is very important to appreciate is that these patients tend to receive care for long periods of time, just because of the nature of that group. In fact, many of these groups are communities themselves because of the large scope of care. This is even though they may change their health coverage and their payers, the plans that they are involved in.

Typically, our members have been early adopters of health-care IT and new technologies. Therefore, the depth and the scope and the longitudinal nature of the data that we have access to through our groups are quite unique.

Most importantly, in doing all of this, whether it's on the group side or on working with the outside

world, using this real-world data, AMGA and Anceta are not only viewed, but are actually acting as a trusted intermediary for our physicians and, of course, their patients.

We have already done a pilot, a couple of years ago, with three groups. We have developed and we are currently operating a production-level data warehouse through a long-term partnership with a company called Convergence CT, using their product suite called DB*FOCUS. Quickly, it loads and transforms the data from a wide variety of sources. It is a technology agnostic, whether it is a practice-management system, hospital, lab, pharmacy, EMR.

Many of our groups, particularly the ones we are starting with, have their own data repositories, where we can extract the data in a semi-standardized form. This securely transports encrypted normalized data into the central data warehouse. Each group holds its own keys to the identity of the physicians, as well as the patients. It is beyond being HIPAA-compliant. We go much further than that in protecting the identities, even to the physicians and the groups at this stage.

Each medical group has access to and control over their data that gets uploaded into the data warehouse.

I am not going to go through this whole diagram.

I don’t have time for that. I just wanted to point out a few things here.

If you look on the left, this essentially represents this staging server, that DB*F, that is located behind the firewall, under the control of each group. That is where the data is extracted and located and pre-transformed to a common data schema. This is pulling from all of their systems. Then we aggregate that data into the centralized data warehouse, where we can form a variety of datasets and data marts for analysis.

On the right-hand side is a suite of activities that they each benefit from for contributing their data. They get access to their own data, actually, directly from the staging server that is before the data warehouse, so they can see what their data looks like in the aggregate, as well as individually. Then we provide routine reports and comparative data, one of our goals. We have a community collaboration portal, where they can not only access the data and receive reports, but interact in a number of ways to rally around the data.

Once they see a study that comes out, they can download a data cube or a data mart in diabetes or cardiovascular or whatever and dig deeper to understand why they are where they are compared to their peers.

Then we have a host of custom analytics, which

could include adverse-event identification, but most importantly, the analysis of that. Some of these adverse events may need to look for patterns, not just that an event occurs and you have a tripwire that you go back and find.

Certainly, we have a number of industry customers that can benefit from this data. That is only fully de-identified, aggregated data that they have access to. This includes not only the pharmaceutical industry, which is shown here, but also payers, other providers, the government, and the like. So we have a system put in place to make that data available, and this is all in establishment with our groups.

The current status:

- We had previously established a very comprehensive HIPAA-compliant privacy practice. This goes beyond just protecting the identity of the patients, but also the providers and the medical groups themselves. There is a business associate and a data-use agreement.
- We have, actually, a common agreement.
- Basically, we have a consensus among those groups. It is part of our role as an intermediary to help them form a consensus.
- We have actually nine participants currently loading three to four years of data, with regular updates.

It is currently monthly. It could be weekly. Some might even do daily through some feeds.

- We have actually aggregated the data among five of the participants.
- We have started, as of last week, analyzing and reporting some of the individual and aggregated data.
- We are currently selecting the next-phase participants, probably 18 to 20 this year, with the idea that we will load from 10.

This is my last slide, except for the thank you. To give you the characteristics of those, right now, among those nine groups, we have about 4 million patients or so, almost 4,000 physicians in 224 facilities, representing 24 hospitals.

This is already a mix of our representation across the country -- integrated health-delivery networks large and small, medical groups, and we are just about to start an academic group as well. All have adopted electronic medical records, some for as long as three to seven years. They are all experienced in clinical research. We are going to be doing 10 to 12 groups a year. So this will go up from 5 million or so this first year to another 10 million or 15 million patients per year, until we have everyone loaded and aggregated.

Thank you very much. Here is my contact

information. I certainly look forward to some further discussions.

DR. SHUREN: Thank you.

We are certainly very interested, in terms of data collection and analysis, as to what may be incentives for those who have access to data to actually pull that data together and perform analyses. I was rather struck, when you were speaking, that what you were setting up was using a standard format from multiple data systems, for multiple purposes -- the data could be used for practice management, for health outcomes, for quality of care -- and then to share those best practices with members.

I wanted to ask you to comment in terms of drivers for getting engaged in this -- one, if some of the incentives for collecting data that then may be used for health-outcomes research or health-services research may be drivers for practice. You talked about practice management. If so, how do you work with those drivers to build this database that may be used for other purposes?

The second is, any thoughts regarding economies of scale? The data can be used for multiple purposes. What has been your experience from that standpoint?

DR. HILL: On the first point, we exist, first and foremost, to serve the needs of the medical groups. It has been a mandate by the board which has organized and

directs Anceta's activities through Anceta's board of directors, most of which come from the AMGA.

Most of the groups can do their own analysis. You are going to be hearing from some of them today. It is not about what they can do themselves as much as how they compare to and learn from their peers. It is the comparative data that is the driving force.

But since they all use different systems, which may or may not -- you could go out and say, "Well, we are going to aggregate the data from this group and that group, and we are going to get the EMR vendors together." But when you have one EMR, you have one EMR. In fact, sometimes when you have one EMR vendor, you have four different datasets within that type of EMR vendor. So someone has to do the heavy lifting in order to combine these among these different technologies. So there is a lot of early heavy lifting on that right side, where we map their data, no matter where it comes from -- whether it is a single site from a data repository or from a practice management and three or four other sources -- to a common data schema, with a hierarchy of coding, not distorting their native coding, but finding some common level of schema and vocabulary, where we can begin to aggregate it and analyze it with some sense. Otherwise, you have apples and oranges. We have apples and apples, but sometimes they

are different kinds of apples.

So that's the tough part, getting it into the form where it can -- we talk about interoperability and standard models and whatnot. That is right on top of all this.

But their real driver is to really have data among themselves so they can see, "Why is my diabetes care not as good as I thought? If it's not as good as I thought, is it because our process within our organization is different than group B's? Or do we have a group of physicians in that organization who are just not quite following where they should be in our own processes?"

You can look at the average of a bell-shaped curve for a measure. Maybe you are a little bit lower, but there is this tail over here of maybe 10 percent of your physicians who are outside of that. Let's bring those back in.

So it's process improvement within the group and among the groups. That is really their driver, improving quality and the process of care -- until more recently, when we are getting into pay-for-performance or performance measures, where we need to have a standard way of helping the payer community understand what quality looks like, let alone pay for. So there is a new motive that we are all trying to figure out. Whether it's PQRI or HEDIS reports

or whatever, it's a daunting task, which has some significant financial ramifications, let alone clinical ramifications. These are businesses.

The second question -- could you remind me?

DR. SHUREN: It was that issue of economies of scale.

DR. HILL: Oh, yes. We have taken a lot of time with the first nine. Halfway through the first nine, we shifted gears, realizing what works best, what does not, in terms of loading, as well as aggregating. We have scrunched the timeframes of loading data significantly, as well as preparing the groups and knowing how to select them better.

This data warehouse is supported by some very large companies that have the big data centers and all of the tools to support the scalability and the flexibility to that.

I think most of the work is done at the individual group level, so the economies are in the process and the learning of the process, as much as that data warehouse growing. However, if you are looking from the user side of the analysis, the larger that gets, the more comprehensive, the larger the populations are, the more you can subcategorize subsets of those populations for a number of studies. So that economy is real.

DR. WOODCOCK: You implied, I think, that as part of the business model, some of the support can come from having the data be mined by others.

DR. HILL: Yes. We make that commercially available to the outside world, third parties.

DR. WOODCOCK: Is that a reasonable source of support for your --

DR. HILL: It's the only source of support. We try to make this as bearable for our members as possible.

We do appreciate the commercial value, as well as the ethical value of having this kind of rich data. That itself improves care. Our groups understand that. We have done it all up front. They certainly are happy doing that. It's always an issue of risk in making our data available. But it only happens because we are doing it through a trusted intermediary. We have two layers of trusted intermediary. We can shut that data flow off at any time if it is not used properly.

DR. WOODCOCK: What you allow is people to analyze on your database? In other words, they would request an analysis, or do you actually give them the data, the anonymized data?

DR. HILL: It can happen both ways. One could order a dataset or a data mart of particular interest, based on selection criteria of patient cohorts, or the

analysis could be done on their behalf by an analytics team that we both are building up, both we and our partners, Convergence CT. So it's pretty flexible.

DR. WOODCOCK: There may be peculiarities of your database that only your own analytics team is able to navigate successfully.

DR. HILL: That's correct.

DR. WOODCOCK: Do you expect that the pay-for-performance, if this becomes more prominent, is also going to be a driver?

DR. HILL: I think so. First of all, understanding what the best method is for that -- we like to think of it as pay-for-results or pay-for-quality rather than process. We have heard people like us say that before.

DR. GROSS: In this process of developing standard formats and terminologies and the like, do you see eventually that the entities that feed the information into this system will ultimately migrate to those standard terminologies? Is there a movement to best practices? Maybe you could speak to what sorts of standard terminologies you use currently.

DR. HILL: First of all, we think so and we hope so, respectively. We have found that there is so much discussion and controversy and variation in what we should

be doing, both from a technical side and from an analysis side, let alone the measures.

We are not trying to introduce anything new. We are trying to look at what makes the most sense among our members of the standard kinds of technologies and measures and coding and whatnot, so that we can help define what works best. We think our groups, as they join, will help to accept and even drive those standards for the rest of the industry, mainly because we can have data showing the value of doing it this way versus that.

As to the coding, it's a little bit over my head. I know some of the words. It's the standard types of things -- ICD-9 and CPT codes. We can HL7 feeds and the like.

As you may appreciate, many of our groups are already doing this themselves. So what we are trying to do is find a common denominator among them that works the best. But even among themselves, they are all a little bit different.

So we are hoping to influence the national standards by showing what works best.

DR. SHUREN: Thank you very much.

We will have opportunities tomorrow, too, to follow up on questions.

Let me ask Dr. Mandl.

Agenda Item: Presentation by Ken Mandl

DR. MANDL: Good morning, everybody.

I am going to talk about informatics and take a step back and think about a framework for approaching a Sentinel Network.

The challenge is that the FDA needs to get innovative drugs and devices to the market quickly, while protecting the public from any adverse effects of those drugs an devices. This is clearly an impossible task. But the perfect is the enemy of the good, so we are going to try anyway.

I come at this from the approach of biosurveillance, an area I have been working in for the past seven or eight years. I would like to just draw some analogies.

In two columns here, we have outbreak detection and drug safety surveillance. Both of these fields were catapulted by crisis precipitants. For outbreak detection, it was anthrax; for drug safety, I think, Vioxx.

The approach traditionally in outbreak detection had been mandatory reporting, and in drug safety, likewise, a reporting system, MedWatch -- actually, a voluntary reporting system.

Limitations: There is bias in reporting. There

are delays in reporting. There is incomplete reporting. There is limited automation and there are limited communication systems. That is outbreak detection. For drug safety, the same. So I think the approaches are potentially alignable.

The action arm has been shoe-leather epidemiology, telephone calls; the action arm here, “Dear Doctor” letters, other public alerts, label changes.

The evolution has been towards automated surveillance systems, as the system I run for the state of Massachusetts. In drug safety, I guess that is the question: What is the next technology we are going to explore to bring this together? Is it going to be an automated surveillance system of sorts, biosurveillance for drugs?

The Sentinel Network, I think, requires four building blocks, broadly speaking: data acquisition, signal detection, adjudication of those signals, and communication of those signals back to the public, the decision makers, the patients, the physicians, the pharmacists.

Let’s just take a quick look through these four building blocks.

Data acquisition: Challenges: Even in these phenomenal databases we will be hearing about today, much

of that information is still in free text. Another challenge: Even things like pharmacy benefit management data, a very high-quality data source for medications, is not necessarily complete. For example, here is some PBM data that we pulled on a real patient and a bunch of drugs he was on. It turns out that four months ago, the patient switched to an insulin pump, which is not dispensed by the usual mechanism. It turns out that the patient was actually lowered to a lower dose of this drug, and that was not reflected, because they didn't get the script change.

Privacy: As much attention as we pay to privacy, it's difficult to actually maintain the privacy of patients in large databases. It is a fairly complex science. We published a paper that actually showed that a bunch of people were re-identifiable from a whole bunch of articles published in the medical literature, because their home addresses were placed on low-resolution maps. It turns out that we could actually put those people right back on their houses, as you see in the lower right-hand corner. So even with the best of intentions, privacy is not always attended to properly.

The opportunities here include processing the free text. There are actual language-processing tools. In the National Center for Biocomputing, an NIH roadmap center, my close colleagues are working on processing

electronic text records from Partners HealthCare to link with genotypes for biological discovery.

Personally controlled health records is an area that I think is important. In work I do with Rich Platt, who you will be hearing from shortly, in the CDC Center of Excellence in Public Health Informatics, we are focused on the public health benefits of acquiring data from patients through personally controlled health records -- essentially, records that the patient owns and uses to integrate their data across multiple sources of care, creating a virtual medical home for patients.

There are models for regional and national data sharing that are evolving. At the Center for Biomedical Informatics at Harvard Medical School, we are exploring these models. One of them we presented to ONC recently. I presented this with Marc Overhage, who you will be hearing from shortly. This is a model that is based on something called SPIN. SPIN is the Shared Pathology Information Network. We are actually sharing data across multiple Harvard hospitals, Indianapolis, and a couple of other sites. I am going to take you through this, in a high-level, conceptual way, very quickly.

Red is the hospital database. Blue is an external database that the hospital maintains, exposed to the network. We get a whole bunch of hospitals or data

sources, with data exposed to the network. The hospitals choose what data they put into this external database. We connect them together through a network. We take one of these guys and make it into what is called a supernode. This guy can actually broadcast queries across to the other sites. What we create here is, essentially, Napster for clinical data, where we have a peer-to-peer exchange of clinical data.

What you can do then is take these sorts of data and share them. I will show you why this is an important model in terms of control and participation.

If you get a whole bunch of networks together, they can talk to each other if these supernodes can talk to each other. So each network can actually have regional variability, and even a different flavor, as long as the networks can interact here through these supernodes.

In the biosurveillance model, we are actually shifting our biosurveillance system to use this network model where the biosurveillance system gets data from one of these supernodes. We demonstrated this, as I said, at ONC across three regions, Massachusetts, Indianapolis, and Mendocino County. We get data out of one of these nodes. This allows control points at two places, which is critical. One is, the institution chooses what it puts into that external database, and two, there are different

routes of access to that external database and you can define authority accordingly.

Here is an example. We could put do biosurveillance initially with anonymized data -- only anonymized data -- exchanged. Then let’s say we have an outbreak or, in the case of pharmacosurveillance, we have a signal of another kind. We can query back, with public health authority, to such a database and actually return appropriate data, based on a prior agreement.

So with these two points of control, we actually decide what is in the database, who gets it, under what circumstances. Those circumstances could include research, IRB approval, public health authority, drug safety surveillance.

Let’s just look briefly now at signal detection.

Challenges: Discovering unexpected drug safety issues requires expansion beyond pure hypothesis testing. However, multiple testing reduces specificity. This is why we can’t have our cake and eat it, too. We can’t do this perfectly. We have ROC curves. We have sensitivity and specificity. So there are new opportunities. There are computational approaches to signal detection. Data-mining approaches can be informed by available data.

In the informatics program, we do work looking at multiple data-source adjustments for multivariate

surveillance. We are looking at new approaches to cluster detection and signal detection. Just to give a reference, there are a number of methods from biosurveillance that I think are highly pertinent here.

Adjudication: This is, I think, the hardest part, because there is always a tradeoff between early detection and false alarms. In this area, if we have an early signal that becomes public, a drug becomes impugned and you end up knocking a pharmaceutical company off the stock market, or at least off kilter. That is a real danger that we need to confront. In order to, I think, engage pharma in this, we need to come up with a system where people feel comfortable about the way these signals are adjudicated. I think engaging Pharma, health care, academia, and government together is critical.

The opportunities here are to provide maximum information for making decisions. I think we need to use open methods so that everyone can see how the result was produced. But we need to be able to create some kind of environment in which those methods are discussed, probably not with a representative of the press there in the initial conversations, so that we can look through these data and have a nimble set of action arms from the Sentinel Network.

At the Center for Biomedical Innovation at MIT, such a safe haven has been created and is, I think,

providing a very nice model for how this might happen.

Lastly, I will just mention communication. Communication, again, is a complex area. A challenge is that it is very difficult to get complete information out to the people who need it in a timely fashion.

The solutions: We can do tailored and targeted information to individuals through informatics, personally controlled health records, point-of-care services. We can leverage experience from a number of areas.

In summary, the three areas for focus, I think, to develop a Sentinel Network, the science that needs to happen:

- One, the computational informatics. That is the clinical intelligence here.
- Two is the industrial-societal interface. I think this is the most difficult one, because there is really no good model for it out there yet that is operational.
- Three is the medical informatics around the data exchange so that we actually have data flowing in the right ways.

I will leave it there.

DR. SHUREN: Thank you.

Let me ask if there are questions.

DR. WOODCOCK: I am very interested in this

question, which I also alluded to, of developing open methodologies that are consensus approaches that have buy-in. I think, if we get enough data, we could impugn every drug in the armamentarium and we could stop using drugs, which would be very safe, but it wouldn't help people very much. So we really need to figure out -- and this is, I think, a new challenge. I would ask all the audience if there is another precedent somewhere in prior human informatics history, or whatever, that we could look at, where we are talking about how to develop, through consensus, methods of analysis that are agreed-upon as reliable. That doesn't mean they are going to get the right answer, but it means that we would agree to go forward with them.

I think, if we can do that, we will be able to get all the partners to engage in this. We share some interests in common in finding out the truth, so to speak, the correct signals and answers.

So I would be interested in your perspective on that.

Also, for any given analysis, say we construct this resource. We have heard about one resource already. Say we construct a larger resource. The question of access to that resource and how that is managed, I think, is another very interesting question that is going to have to

be discussed.

I would like your comments.

DR. MANDL: I really think this is the hard part. In the outbreak detection world, we started worrying about the same issue. What we worried about was, when these things went off and signaled that there had been an anthrax attack, how would we inform society that there had been an anthrax attack? What would be the tradeoffs to an early warning, with panic and stock exchange. It turned out that in seven years of this, none of these systems has signaled an anthrax attack. We kind of forgot that that was what we were initially worried about in these sorts of decision frameworks.

But I think that we are not going to have that kind of honeymoon period here in drug safety surveillance. Quite the opposite. The interesting thing about sharing this data is that the data really will need to be available so that the analysis on which decisions are made can be replicated, and the methods need to be open so that they can be replicated. I think replicating the same finding across multiple datasets and having a multiple-modality approach to following up on early signals, and one that is rapid, is going to be the key. So having that nimble -- and I think there are probably a lot of individuals in the room who represent just such organizations and are stewards

of such sources of data and sets of skills for analyzing the data, who could become part of a network that rapidly responds to an initial signal.

Then we have to understand the value of data in different datasets and by different methods, in order to know when we are starting to reach something that represents a preponderance of evidence.

DR. SHUREN: Other questions?

MS. CRONIN: You mentioned the NHIN and the regional emerging models for data exchange. In the work that you have done so far, have you started to identify the requirements or the data models for drug surveillance, in particular?

DR. MANDL: I think to get started in this, it would be some of the standard clinical data that we need to identify adverse events, plus medications. The more detailed information we have about medications, the better off we will be in terms of being able to truly define that exposure.

But I really think it's more the analytics about how we use those data than the data. I think the data are, on the whole, the usual suspects. Then really being able to define exposure and outcome is another set of tricks.

MS. CRONIN: I'm just thinking, in terms of developing the infrastructure that will be evolving over

the next year, particularly as the next round of the NHIN gets funded and it does have more of a regional and state orientation, it would be helpful to know if there has been already some thinking and some work done on what the requirements are, whether you need a hybrid approach, whether you need some kind of repository so that you can have easy access to data that you need to do this type of analysis. Or could it be something completely federated, where you are just going to query on demand?

That is an example, but there are a lot of other issues that would need to be resolved, I think, to make this work on a regional or statewide basis.

I am just curious to know whether or not anyone in CBI or maybe in Regenstrief has thought through that a little bit.

DR. MANDL: I'm sure Regenstrief has thought about this. I'm sure Marc will have an opportunity to respond as well.

I think it is a hybrid approach. Some data will need to be put together so that you have a broad sweep.

On the other hand, I think there are individuals here representing very large organizations in which data could be analyzed locally and signals then combined subsequently.

So I think it will be some combination of

combining data and combining signals that we need to develop. A lot of that is going to be beyond what would be ideal, but what would be practical from a sociopolitical perspective. I get the feeling the data will tend to stay where it is, no matter what we do, and so we should definitely conceive of a hybrid approach.

DR. BRAUN: One thing that you said struck me. I first want to make sure I heard it correctly. I think it was with respect to biosurveillance, where you were saying that the sites that were in the periphery would send in the information that they wanted to. You did say that? Okay.

In population-based reporting systems, my understanding is that they operate under the understanding and the standard that there will be complete reporting and you will know denominators and numerators. Can you explain how analysis of such data, where you would have heterogeneity of what is submitted, would be accounted for?

DR. MANDL: In this model you would have a core dataset that would be a minimum dataset. But then you might contribute more, let's say. For example, in biosurveillance, perhaps some institutions would contribute detailed location data -- perhaps some would be at the level of the zip code, some at the level of the county -- and then you would deal with what you had when you got there. But you would have the widespread participation at

that level, rather than choosing just the least common denominator and posing that widely. It's a way to get least common denominator plus additional information from different sites.

DR. BRAUN: Certainly, then, with the common dataset, the advantages are clear. The analytic complexity of dealing with heterogeneous submissions from large numbers of sites is certainly something that impresses me. That is just an observation.

DR. MANDL: Absolutely -- and depending on the precise application for the data. We are talking about repurposing networks for multiple causes. Let's just say that clinical care is cause number one. With clinical care, the more, the better, but we will take what we can get. That's clearly the way it is. Right now what we have is nothing, usually, when a patient shows up. So anything is good.

So the system is designed to think about all these different opportunities.

You are absolutely right. For certain research purposes, you really may want to define a common dataset across those sites. But by having the same network making data available for clinical care, for research, for public health surveillance, I think the institutional control over what goes in will obviate the need for consensus building,

which doesn't usually work very well.

DR. SHUREN: Thank you very much.

Dr. Overhage?

Agenda Item: Presentation by Marc Overhage

DR. OVERHAGE: Thank you very much.

You can tell who didn't get their presentation in on time. Ken is always two steps ahead of me. Thank you for setting things up so nicely, Ken.

I am going to share a little bit about the work that we have been doing over the last 15 or 20 years to build a regional population-based health-information exchange, and then a little bit of the experience that we have had in leveraging that, primarily for using it as a tool for measuring outcomes in various kinds of clinical studies that we have been doing.

This infrastructure that we have built, which we have been calling for a while the Indiana Network for Patient Care, is in many ways comparable to the kind of infrastructure that Ken was describing in a local market. The market that we are most deeply penetrated in today, just to give you a sense of scale, is this: It's about 120 miles square, about 4,000 physicians, about 1.9 million people who live there today, although, obviously, data are captured longitudinally, and so we have data for about 3.5 million people in the repositories today.

The way that we attack this problem -- and Ken talked a little bit about leaving the data where they are -- one of the challenges that we find is that, leaving data where they are, you run into all kinds of little "gotchas." For example, laboratories keep data for a very short of time, typically -- 30 days, 60 days, something like that. Other systems roll things into offline third-tier/third-world storage that you can ostensibly retrieve it from, but you may or may not be able to.

One of the challenges is that, even when you do that, when you centrally manage those data or you leave them where they are, there are some core challenges you have to address, including how you link together the patient's data. There is not a common identifier. The PBM uses one descriptor for the data; the health plan uses another; the physician practice; the three hospitals they have gone to. Just think for a moment about, for example, the Medicare population, who sees, on average, six physicians a year. They are not all in one place.

So one of the first challenges is how to link together patient data. Every institution typically has some kind of unique identifier for the individual patient, associated with a variety of demographic data -- date of birth, Social Security number, things of that nature. That unique identifier, unfortunately, isn't linked to anything

else.

What we do is, we take advantage of the fact that everybody has a relatively common set of demographic data. Currently we use a fuzzy deterministic algorithm to link together those unique identifiers across multiple institutions. We have examined a variety of other strategies as well. It turns out you can do very, very well.

I should point out that even if you have a common national identifier, it doesn't solve your problem, because those are erroneous a substantial number of times -- the same kinds of problems with Social Security entry or anything else. In countries that do have a national identifier, it is often not used as the identifier. For example, in Great Britain, if you go and try to retrieve the patient's data using their National Health identifier, it is often not a searchable field in the database, in the practice-management system. It is recorded as an attribute of the patient. And there are the usual problems of people who don't have one, who have multiples, and so on. So you still need these kinds of strategies, even if you have a better national identifier.

We use those demographics to deterministically link together those different identifiers for the individual.

The second challenge that we have is that the data are simply not identified in a common way. A serum sodium at every one of the laboratories in this country is called something different.

Unfortunately, we are not so clever in solving this problem. We use the example here of immunizations. Different organizations have a different way to identify immunizations. There are national codes for these things -- in this example, LOINC codes. Unfortunately, organizations don't use them, by and large. There is a whole variety of reasons for that. It is going to be quite a while, I believe, before those are adopted and driven into the operational systems, like laboratory systems, radiology systems, and so on. There is a whole variety of reasons behind that.

So we take these standardized national codes and map them -- and, yes, that is a lot of work -- so that we know that a serum sodium from each of the different services is LOINC code whatever-it-is. (Contrary to public opinion, everybody at Regenstrief hasn't memorized the entire LOINC dictionary of 40,000 terms, so I don't know what it is.)

The other thing I will say about this is that people often forget that physicians also have multiple identifiers. I have now 49 unique identifiers as a

physician in our market, including my national provider ID. Yes, I am one of the 1.3 million who has registered so far. But I have 46 other ones, as well, that are in common use. We use there a human-mediated process to link together. In other words, we have a person who sits down with the physician and their practice-management team and says, “Yes, that’s him. Yes, that’s him. Yes, that’s him,” typically because there is not enough metadata to combine that together.

If you go through this process over and over again for a variety of different health-care participants, including payers and others, we start to create -- Wes Rishel coined a term that I like, “edge proxies” -- in other words, these databases of the nature that Ken described that represent the data from this individual institution, separated and segregated, and in our case centrally managed, for a variety of reasons, with that global patient index, a global provider index, and a concept dictionary that link that information together.

One of the other things that we do is -- a great deal of information is still locked up in text reports. Even if a physician has an EMR, many of them use it as a glorified typewriter, and the information that is retrievable from that system is a blob of text. So we incorporate various levels of -- I put it in quotes,

because I hate to glorify it by calling it “natural language processing.” We use that set of tools to extract key concepts, including, very importantly, negation -- in other words, “The patient denied nausea and vomiting,” is a very common kind of expression to find -- and store those back into the repositories as structured data so that they can be used.

Basically, we capture data in real time from data sources -- laboratories, radiology, PBMs, things of that nature -- clean them up, if you will, and then we drive a variety of processes off of that. So the Health Information Exchange really is that common flow of data, structured in a consistent fashion so it can be used for a whole variety of services.

As Dr. Hill described earlier, it’s critically important that you do that, because you need to use these data for more than one thing if you want to be able to pay the bill at the end of the day. These are expensive processes, to normalize these data, to build this infrastructure, to capture things. If you can’t find multiple income streams, you are going to have a hard time funding it at the end of the day.

But very importantly, in the middle of that is a concept that we call “negotiated access.” Everybody owns their data. We subscribe to the principle that the patient

owns the data, but people are custodians of the data. They all want to make sure they know exactly what is going to happen to their data, when, by whom, and I think, going down the road, as these data are used for commercial purposes, how much am I going to be paid?

We have used this infrastructure, as I said, primarily to look at outcomes and research projects for adverse drug events. We essentially do that with a batch process that each day scans this federated repository with a distributed query and identifies candidate or potential adverse drug events. But we have found that for study purposes anyway, we have to incorporate a human review process in order to really determine whether those are true adverse events or not, for the usual reasons that you run into. The fact that somebody died while they were taking a drug obviously doesn't mean it was due to the drug, but they did. You can build various levels of intelligence in to refine that so that the positive predictive values get up into the range of .2 or so -- in other words, that 2 out of every 10 represent an actual event. That is about the best that we have been able to do to date.

We incorporate a process of tiered review, where the computer identifies events from the database; data analysts -- not clinical people -- are able to screen about 90 percent of those out. They are able to say, "Nope, it

ain’t an adverse drug event,” based on some tools and algorithms that we have developed. Only a minority require clinical review. We only do this when we are doing projects. The reason we do that is that even this approach, which is fairly cost-effective compared to traditional approaches, costs $40.00 for every ADE we find.

So that is about where we are today in our ability to do that.

As I say, for a variety of studies we have done -- these are some data from a collaborative study we did with the folks at Brigham and Women’s Hospital, using this same approach, same set of rules for identifying ADEs in the two markets, and over the course of a four-month period, identified a variety of ADEs, with, as I said, a positive predictive value of around .2.

The last thing I want to close with -- and Ken, fortunately, set this up, so I don’t have to explain it all -- is that we have been leveraging this SPIN tool, the Shared Pathology Information Network tool that Ken alluded to, which was created under an NCI contract, as a tool for discovery, for data mining, if you will, that allows us to do exploratory queries in a completely anonymized fashion across that very large distributed database. As an example of this, a few years ago, one of our fellows at the time heard at a conference a case report that suggested that

erythromycin might be contributing to the incidence of pyloric stenosis in infants. It is fairly common for infants to get erythromycin for a variety of reasons. He came back and, within 48 hours, was able, using this tool, to very convincingly demonstrate that, yes, indeed, there was an extremely strong association. I think that has been taken forward into common practice today.

That was the last thing I wanted to share. I look forward to the conversation throughout the day.

Thanks, Ken, for saving me several minutes of discussion and keeping me on time.

DR. SHUREN: Thank you. I like that sort of economic approach of trading off time between speakers.

Let me ask if there are any questions from the panelists.

DR. GROSS: I have a question. This pertains to the last speaker as well. Don't forget about medical devices. To what extent are you thinking about medical devices?

DR. OVERHAGE: It's a very important question. Given where we live, with Cook and most of the orthopedic devices in the nation manufactured in the northern part of the state, I am very sensitive to devices.

Devices are generally poorly captured in routine clinical care today. We have been working, but have not

gotten very far down the road, on focusing on capturing the details of the individual device more reliably in routine clinical care processes, not just for purposes like this, but also for purposes focused around economic analysis and a whole variety of things -- the data reuse issues that Dr. Hill alluded to and I am sure you will hear about the rest of the day.

Certain things -- for example, implantable cardiac devices -- we are actually pretty good at now. I can tell you if you have a drug-eluting stent, of what brand and so on, fairly well across that patient population. But we have a long way to go.

DR. GROSS: Just briefly, how did you get to that point with ICDs?

DR. OVERHAGE: We didn't rely on ICDs to do that. Instead of relying on the billing codes, we capture data from clinical information systems. For example, these particular data we focused on capturing through cath lab systems. There is often a way to record this kind of detailed information in the clinical care system, but people don't always do it.

So building the process and the value case for capturing that at the point of care has been the challenge. As I said, for cardiology that is going well. For orthopedics, we have a way to go yet. We have lots of

other things on the list.

DR. SHUREN: Let me just ask one follow-up on that. Whereas for drugs we have a national drug code, we do not have a national system for device identifiers, currently. How do you wind up, then, identifying devices? Do you have some internal codes being used or some other way of capturing that information?

DR. OVERHAGE: It's a very good question and opens a wonderful door that I will take advantage of.

The question about the device codes -- because we have taken a fairly targeted approach, initially, and focused on some of the cardiac devices, we relied on manufacturers' model numbers as a fairly simple way to represent them. They obviously needs to be a more robust way to deal with that.

The door that you opened that I will take advantage of is, NDC codes stink. I think anybody who has worked with them will tell you that one of the big challenges is that 10 percent of the NDC codes we see in the wild are not real. In other words, if you take streams of data from clinical systems, 10 percent of the time you don't know what it is. Second, for much of what we do, it doesn't aggregate things at a helpful level.

The good news is that I think the very strong work that the NLM and FDA and others have been doing on

RxNorm is coming a long way in creating tools that will facilitate that.

I appreciate you opening that door, because it's one of my soapbox issues.

DR. WOODCOCK: May I just add something to that? What we are doing on drug registration and listing also affects the sort of bogus NDC codes, but it won't affect the aggregation, which the medication terminology is aimed at fixing. So we are trying to work on that.

My question: You said that local groups are going to want to continue to have control over their data. Does that mean per analysis? If you were to participate in a national analysis of some adverse event, that would be on a per-analysis basis that you would like to have that control?

DR. OVERHAGE: That is a very good question. In other words, we have this wonderful store of data now. How free can you be with it? Again, we always treat them as the participant's data.

The way that we have managed that -- and I don't know if this is the answer -- there is a local management group with broad representation that we bring categorical uses to, not individual uses. For example, public health surveillance is a use case, if you will, that we brought to the group, and they said, "Sure, under the following terms,

you can do anything that you need to do for public health, as long as you don't cross this line." For research purposes, as long as it is anonymized and so on -- you don't cross that line.

There are things that they get real nervous about, and so there are things that get reviewed. But by and large, the process has been, single IRB approval, if IRB approval is appropriate, single point of review for categories of uses, and then as new and novel uses are created, we stimulate a lot of discussion and dialogue.

MS. CRONIN: Marc, I just want to thank you for what I think is your 10th presentation at an HHS public meeting this year. It was very helpful.

I am wondering, since you have had a lot of experience in working with industry and other academic partners in doing this kind of work, and you also are very familiar with everything happening in ONC, how do you think this whole area could be advanced? As we know, the NHIN is going to be rolling out. What is the best way for FDA and academia and a lot of these regional initiatives to be working together over the next couple of years?

DR. OVERHAGE: It is obviously a big question. If I knew the answer to that, we could all move a lot faster. It is not a simple answer, I don't think.

I think some of the key things are -- and Ken set

up a number of them very well -- and I think drug safety is a wonderful example for working this through: How do you take a resource, an information resource, which certain people feel like they own or it has value, and bring that to the public good? Those kinds of policy sorts of issues, I think, are the one of the great challenges.

The second area that I think needs a lot of attention, as Ken described, is the analytics. I didn't talk about that at all, but I think that is an extremely challenging area that we have just scratched the surface on. Actually, some of the dialogue earlier -- it's easy to do analytics when you kind of control everything and it's all laid out nicely. When you accept the messiness and ugliness of the real world, the analytics get a lot uglier and messier.

I think those are the two developmental areas, in parallel to all the other things that have to happen in terms of moving data standardization and so on forward that there are many initiatives around the country already trying to move.

That was a lousy answer. Hopefully, by the end of this meeting, I will have a much better answer.

DR. SHUREN: Thank you very much.

Dr. Platt?

are the prime target for this vaccine, to be able to study that.

We are embarking on a health-plan study that would be what epidemiologists would call a plain-vanilla cohort study in four plans that have 40 million members. I think Miles Braun was exactly right in saying that you need to have defined populations. By their nature, health plans serve defined populations. This study will use linked claims to identify individuals who are immunized, to sort out post-immunization person-time from other time, to identify potential cases, and then do medical-record review to confirm those potential cases. This is a study that has been vetted with FDA and CDC. There are clear rules for making interim reports to the regulatory agencies and a public report at the end.

On the basis of that kind of experience, the CERTs are fairly well along in planning to develop a health-plan consortium for public health. Its goals will be to improve on the safe use of marketed vaccines and prescription drugs. The target population is 100 million individuals. This will be an activity of the CERTs, which I described very briefly while the slides were coming up.

The aims are to address both the prospective signal detection for therapeutics and to be able to do detailed assessments to follow up specific questions that

I am going to stop talking about Leapfrog and I am going to address a couple of things that were the questions, if that's okay.

In addition to working for Leapfrog, I also do consulting for the National Association of Health Data Organizations, and I am a former bureaucrat in a state. I ran a health data agency, where we collected 26 different population-based data systems, including vital records. Actually, no one has mentioned that today as a possible source for information, but death certificates certainly are a potential source for information.

I also worked at the University of Wisconsin-Madison, where we have in our research center the MDS, which is the Minimum Data Set for nursing homes, which is by no means "minimum." It has lots of clinical data in it.

I want to echo some of the things that were said about data systems that are out there that we might make use of, from my experience. Medicare data was mentioned, Medicaid data. The MDS is a potential, death certificate information, some of the registries that are funded by the CDC and others in states, state discharge systems. There are now 38 states that have hospital discharge data available. It is put to a lot of different uses. In many cases, it is available as public-use datasets.

There are issues, obviously, with any type of

large data system. There are issues related to the codes that are available. There are issues related to the privacy of the data.

There are issues related to the staffing, which was brought up, too. In order to work with these very large data systems, you really have to have a really good understanding of the data. You don't get that over multiple data systems. You get that by working with one data system for a really long time. Yet what is really important is integration. But there are lots of barriers to integration, one being that most individuals working in data analysis and collection are experts in one area, in one data system. When you begin to cross over, you have some issues, because you don't really understand the data well enough as an analyst to fully appreciate some of the nuances of the data.

You also have proprietary issues. No one has mentioned this. If I collect a particular type of data, I am the owner of that data, whether it is a discharge system, whether it is a death certificate system. Getting those individuals who are data collectors to actually share data is no small task. In fact, I have done a white paper through NAHDO on the barriers to data integration across systems. This is not an easy thing to do.

There are some examples where it has worked.

It affects patient care.

Thirdly, if you have clinically relevant treatment advice, including safety advice, and you provide it in a user-friendly format at point of care, health-care professionals are very receptive.

In summary, there is a lot of good data to suggest that the clinical community is receptive to this type of information. They must be part of the equation, because they are the ones seeing the patients, they are the ones entering the data, and we have to improve their knowledge of how this data is being utilized. It has to be an ongoing, sustained process. Frankly, there has to be money involved, because these programs cost money.

Thank you. I will take any questions.

DR. SHUREN: Thank you, Steve.

My apologies to the past speakers and Dr. Goldman for my having to step out and miss their presentations.

With that, let me ask the panelists if folks have any questions.

[No response]

You answered everything.

DR. GOLDMAN: I presume he wants to ask a question. Does he get a chance to ask a question?

DR. RUGGIERI: I think you may have misunderstood me. I apologize. I was speaking very quickly. In no way

do I think health-care practitioners should be left out of the, quote/unquote, equation. Health-care practitioners are involved and have a lot of their day absorbed in capturing the essential data that we need to discover information about risk associated with products.

What health-care providers should not be burdened with is to step out of their exam rooms and away from the patients and have to worry about filling out a burdensome form or a burdensome set of information to relay that information. Plus, one health-care provider's hypothesis about a relationship of a particular event with a drug is exactly that; it's a hypothesis.

I think what this system is that we are discussing here today -- I am not saying that there should not be a mechanism for physicians to call to attention of authorities their concerns about a particular product. There should be an AERs system. I am not asking that that be replaced. But I do think we need something broader. I think we need to be able to keep our ear to the door and watch not, quote/unquote, adverse events, but watch all events. I think the knowledge about risk is going to come out of watching patterns, not individual events.

DR. GOLDMAN: May I say something? Thank you for the clarification. Let me make two points.

Given that the meat is going to be infrequent

However, they are obviously giving up some control of it, in that we are aggregating it at our site. We really had to go very far to assure them that the risk to them was low. We had to be HIPAA-compliant, and really even more than HIPAA-compliant. They are business partners with us, but we made the decision to try to have as little protected health information at our site as possible. That made them feel better, as well.

So we remove all of the usual patient-identifying information, strip out name, Social Security number, phone, address. You can see the rest there that we do. At our site, it is all de-identified and stripped of protected health information. They are able to re-identify the patients at their sites. If they are running a quality report on our site and want to dive in and see which patients are not meeting their A1C goals, they can actually look up those patients on their own systems.

The type of information we have is the usual types of clinical-list items:

- Problem lists with ICD-9 codes. We have more granular, kind of proprietary vocabulary, but we found that most people are comfortable with ICD-9s and are using ICD-9s. We would like to get to SNOMED at some point as well.
- Medication data. This is all mapped to a commercial drug database vendor, with their coding system,

as well. That is coded information.

- Flow sheets. We have the usual vital signs, lab results, and even some social history types of things, like smoking status. So there is a richness there.
- As I mentioned, the demographics that we do have.

The thing that we don't have right now is, we don't bring the document information to our database. Again, that was due to concerns about protected health information. We don't even bring that in.

Just some quick statistics. We have over 100 members. A member can be anything from a one-doc office to a large academic medical center. We have over 6 million patients -- I think it's getting closer to 7 million now -- 6,000 physicians, mostly primary care, but we do have specialty information as well.

Again, we have had this in place for about five years. I think somebody else's slide showed that the average length in the database is at least 22 months. With a clinical dataset, that is actually a little harder to define. Due to time constraints, I won't go into that.

Just some of the considerations about this:

- There is always the question of representativeness of the data. We do have certain regions of the country where we are more highly penetrated and

DR. SHUREN: Thank you.

Next I would like to call John Rothman, Opt-e-scrip.

Agenda Item: Presentation by John Rothman

DR. ROTHMAN: Thank you.

I am here to present to you today a collaboration between Opt-e-scrip and the American Academy of Family Practice. The first half of this presentation was supposed to be given by Wilson Pace. Wilson is not here today. Actually, two other people were supposed to give this presentation, and they both called me at the end of last week, and I have both of their slides. So I have to do the best I can here.

This is a collaboration based around a novel way of prescribing chronic-care drugs developed by Opt-e-scrip and currently being implemented by the American Academy of Family Practice. The AAFP is a very large and successful practice-based research network. It has a large number of physicians, in virtually every state. They have a core staff of 15 people. They have numerous peripheral networks accessory to their central network. They are actively involved every day in practice-based research. They have numerous protocols going on. Opt-e-scrip is working with them currently in the development of this chronic-care methodology.

Currently, this method is being looked at by AAFP as the standard of care. We are assessing it right now in a 600-patient trial with AAFP. Other trials have been written and other trials are going to be fielded soon.

There is a lot of literature on this if you care to investigate it. The American Medical Association came out in favor of designs like this.

These are the kits. The drugs are blister-packed. You can see that they are blinded.

This is the regimen in which patients are given different drugs on different days, in multiple-crossover design. Again, that is the kit.

This the data-collection form. This is a week's worth of data. Two sides of that one form are a week's worth of data. So it is very abbreviated, very quick to collect, but it does hit the major points on safety and efficacy, and it does give you a profile for the patient and their ability to both respond to and tolerate a given drug -- for example, an H2-blocker versus a PPI.

This is the cover page of the report. It tells you some of the statistics that are used. There are some of the statistical comparisons between the two drugs within a given patient.

As I said, this method is being used currently by AAFP.

That is my presentation.

DR. SHUREN: Questions?

DR. TRONTELL: Just looking at your slides, it would suggest you are looking largely at outcomes that are symptomatic, where the patients can report. Is that a fair assumption?

DR. ROTHMAN: Yes, that is a fair assumption. This is intentionally designed to be very real-world. It is taken in the context of which drug is optimal for any given patient in the universe of potentially optimal drugs, and to narrow that choice down based upon evidence. This is an evidence-based way of prescribing specific medications.

DR. TRONTELL: How would you see this being applied to other instances where, perhaps, you may not have such ready measures of effectiveness, where it might involve physician monitoring, laboratory monitoring, et cetera?

DR. ROTHMAN: There is a variety of applications, and we are beginning to explore a number of them now. One of them, for example, is in the area of ADHD, where one of the characteristics of the disease may, in fact, be a response to the drug overall. It can be used almost diagnostically in that setting.

So we are looking at a variety of things.

Effective communication in a medical products safety data system involves dialogue among all stakeholders. The deficiencies in existing data-collection systems raise issues regarding when to notify, what to say to whom, and how patients and providers will be affected. Successful communication requires addressing these deficiencies.

Graphical methods for expressing such concepts as relative risk and the magnitude of rare events can help provide a perspective on risk that facilitates evidence-based decision making. Involvement of health practitioners at this stage is also essential.

State-of-the-art technology and processing techniques should be employed to provide secure, efficient, and transparent experiences. Extensive validation and error checking will be needed to produce high-quality data for risk evaluations. Customized and intuitive graphical user interfaces will maximize the accessibility and usability of data.

Thank you very much for your time.

DR. SHUREN: Any questions?

[No response]

Thank you very much.

Ed Helton, CDISC.

Agenda Item: Presentation by Ed Helton

If you are an MD, you can gain access to the system by answering a few challenge questions, and we will verify who you are. But then you can come into the system, look at the data, and participate.

DR. BRAUN: You said there were 10,000 physicians. They are distributed among a wide variety of specialties? Is that fair to say?

MR. FROST: That's correct. Certainly at this point, we have physicians in all states, everything from tertiary hospital settings to individual, very small practices, and across all specialties. We do not burden the physician-participants with putting in information on their demographic background. However, in a de-identified way, we can couple all the data created in Sermo with pretty substantial information about the specialty, place of practice, and other demographic information about the practicing physicians.

DR. BRAUN: Just two more quick ones. How frequently do they visit, on average?

MR. FROST: One of the constructs I described in this -- well, I will try to pull things from real data. We had 365 new physicians register just yesterday. One thing that is very substantial in this is that Sermo really can be grouped with Web 2.0 phenomena. I don't think there is a very close alliance, but there is some alliance in this.

This concept is to use online communities to present information and filter information.

Typically, online communities are composed a very high percentage of people who are just observers, lurkers, and a very, very small percentage of people who are participating. There can be much less than a fraction of a percent of the community itself participating.

By creating a very simple quantitative mechanism, a survey -- “have you seen this case under these circumstances,” or, “do you think that there is a causal relationship between drug X and this syndrome or symptom” -- we have user participation rates of 15 percent and higher. That is part of the structure and the simple data object -- so very, very frequent participation.

DR. BRAUN: Can you briefly define, for those of us who don’t know, what prediction markets and information arbitrage are?

MR. FROST: Sure. The concept here -- fundamentals. A prediction market is theory and application from the economics field. The fundamental premise of prediction markets is that you can aggregate information from a group of people, in which no one knows the answer to a particular question, but if they have diverse enough backgrounds, you can aggregate that information effectively to come up with the answer.

presenters. All of those will be made available on our Web site by close of business this Friday. It's the FDA address, www.fda.gov/oc/op/sentinel/. It will be posted up there by close of business Friday.

Anyway, thank you again. We will see you tomorrow.

(Thereupon, at 4:30 p.m., the meeting was adjourned, to reconvene March 8, 2007.)