- [Rebecca] Good afternoon, everyone. This is Rebecca Wilgus, from the Duke Clinical Research Institute, and I'm here with Dr. Jimmy Tcheng and Asba Tasneem, and we have others on the line. We'll be getting started with our webinar momentarily.

- [Dr. Tcheng] Hello?

- [Rebecca] Hi, everyone. This is Rebecca Wilgus, from the Duke Clinical Research Institute, and we'll be going ahead, we're ready to get started with our webinar. For everyone's benefit, that was not able to join us today, we're gonna record the webinar. Let me ask at this time, if you're not speaking, if you could please put your phones on mute. We're getting some background noise, and it'll help our presentation come through more clearly, if everyone will please mute their lines, if they're not speaking. When you have questions, if you'll please use the chat pod that's located in the bottom of your screen. And if you'll make sure that you chat your questions to everyone, so that we can all benefit from those questions, and learn together. At this time, I'll turn it over to Dr. Jimmy Tcheng.

- [Dr. Tcheng] Thank you, Becky, and good afternoon, or if you're further west of us, good morning, and welcome to today's webinar, on improving healthcare data interoperability. This is the third of a series of meetings that we're having in support of the project to try to improve healthcare data interoperability, particularly from the focus, from the vantage point of, from the vantage point of clinical registries. Today's presentation will largely focus on our findings, and the recommendations that we've derived from the work that we've done so far. The outline of our presentation this morning, or this afternoon, as the case may be, will be to do the following. One, is we would like to provide you a brief overview of where we are with the project. That will be followed by a review of the core common data elements, that will be developed as the final work product of this particular grant. There are two components to that, the first is fairly straightforward, and that's to review the metadata concepts that we are developing for these core common data elements. And I use the word core carefully here, because others have used the term core to indicate a set of frequently used, almost universal data concepts, specifically the USCDI. I'm gonna be using the word core here with the lower case C, specific to our presentation. Most of the time will then be spent on review of the CDEs themselves, we'll be going through them in fairly rapid-fire fashion. The reason to do this, is to make sure you understand how we've gotten to the recommendations that we're providing, and to arm you to do the homework task that we will be asking you to accomplish, at the end of the presentation, and that is to take the spreadsheet, if you will, the overview of all the content, and review it in some detail, and return us comments. So with that, if I could remind everybody on the phone, please, put yourself on mute. That will help everybody with listening to the presentation here. Just a handful of slides, to talk about the problem itself, and the project. I believe most of you, but perhaps not all of you, have participated in the series of meetings that we've had so far. So just to try to level set, what we are trying to accomplish
here, is achieving true interoperability, not from the standpoint of the interoperability, or the
movement of, say for example, PDF document-based information from one clinical context to the next.
But instead, setting the groundwork, so that certain specific, and again, we'll call this core clinical
concepts, are captured at the point of care in a way that then becomes universally interoperable, with
very little, or no transformation of that information, as you go from one electronic health information
system, to the next. Why are we focusing on the registry community? Well, the registry community
probably is one of the better places, at least I believe, or we believe, it's one of the better places where
this work can actually be accomplished. That the work of the physical work of developing databases that
are interoperability-ready, will best be exercised. Our project goals are the following. The first one is, to
evaluate the current state of registries. Specifically, to identify the concepts that are shared across the
registries, and then to assess the compliance, if you will, of registries, registry vendors, and developers
in the use of data standards of those, for those concepts. We also wanted to take a look at the predicate
work in the space of data standards of those, for those concepts. We also wanted to take a look at the predicates
in the space of data standards of those, for those concepts. Our project goals are the following. The first one is, to
evaluate the current state of registries. Specifically, to identify the concepts that are shared across the
registries, and then to assess the compliance, if you will, of registries, registry vendors, and developers
in the use of data standards of those, for those concepts. We also wanted to take a look at the predicate
work in the space of data standards of those, for those concepts. And then, the most
important thing, is to create an implementation guide, and that's a package for database developers, of
recommendations, so that if the data elements themselves were developed in compliance, if you will,
with the implementation guide, you would be ready, or near-ready for interoperability with a little, with
a minimum of fuss, or loss. I do need to acknowledge, again, the participating registries. I hope your
name is up here. If you are participating on this call, and you have not contributed your content, you're
still more than welcome to be on this call. And please, again, accept our thanks, for all those who have
contributed content. The progress report is as follows. We had three aims that we created for the grant,
that we supplied, or that we submitted, I should say, to the Pew Charitable Trust. The first one was to
perform a Case Report Form review. Originally, we had actually said, that we would just review about 20
of them. Given the wide response from across the registry community, we ended up with content from
38 different registries. That Case Report Form has been reviewed, and we created what are called
concordance analyses, that is, how frequently a concept appeared in one registry, to the next. It didn't
have to be exactly the same data element, the important thing was the concept. And then, that was
used to help develop a list of CDEs, that we've further developed, as part of this project. Aim number
two, was to do the same exercise with national data standards, specifically, national data models, like
PCORNet, the OHDSI data framework, etc. Again, we conducted a concordance analysis, which also, then
contributed and informed the development of the CDE list. Aim three, is where we are today, and that's
to discuss our findings, specifically, for the CDE recommendations, in terms of what we would consider
to be now, again, based upon our surveys, and based upon bringing together every bit of information
that we could find across various data models, et cetera, into an implementation guide. Our anticipation
is that we will have this ready on or about December 1st. Originally, we had set a date of October 1st,
which is only a few days away, but in thinking about this, we actually did want, we want to solicit your
opinion about this. And so, again, at the end of this presentation, you're gonna have a bit of a
homework assignment. That's a major part of the reason for extending the deadline, is because of the
real need to make sure that we have representation of the community that's attending the webinar
today. We did start from a point of having at least a bit of anticipation of what we would see fairly
commonly across the registries. This has to do with my participation, particularly, in the American
College of Cardiology's NCDR registries, the National Cardiovascular Data Registries. These core concepts
of interest are listed here. I'm not gonna read through them, you can read through them yourselves.
But, suffice it to say, this our starting point. What we anticipated, was that we would find that across the
registries that we surveyed, we would have approximately two thirds to three quarters of the registries representing combinations and permutations of the core concepts of interest that you see here. As it’s turned out, that was a little bit naive, it is a bit less, it actually is quite a bit less than that, in terms of the commonalities for some things. For example, smoking status, even vital signs are not represented all that frequently, not as frequently as we would expect. But we have elected not to discard those things, because, they do appear frequently enough that we felt that they should be developed. The second point that we brought to the table, was a fair amount of work in the CDE development arena, dating back some 10 years. We have had formal training and work in the representation of CDEs, per the ISO 11179-3 standard. We’ve worked in the past, for example, with the National Cancer Institute, EVS/Enterprise Vocabulary Server, and the systems that kind of out of that. The NINDS, the National Institutes for Neurologic Disease and their data repository, their CDE data repository, etc. And from that, we’ve brought forward these concepts, tweaked just a little bit with contributions from the Women’s Health Technology Group, a coordinated registry network that’s being setup by FDA, as well as, other entities, that have led to this list of CDE Metadata. I think this is actually an important slide. I do want to spend just another second talking about these concepts, because, these are the things that you will see represented on the content slides that are to follow, for the CDEs. So what we have as the metadata, is: number one, a concept label, a clinical concept label. You can think of this as a human prompt that would be on a case report form, or on a data entry screen. A clinical definition of that particular concept, and when I say definition, I am not talking about instructions on how to complete the form, such as if you have a checkbox for hypertension. What I did not want to have, is a definition that says something along the lines of check this box, if the patient has hypertension. No. What we aimed for, was actual definitions of the terms, of the concepts that you’ll be seeing here. Where there is a restricted list of allowed values, that allowed values list has been represented both in a human form, in a human prompt, if you will, for again, the case report form, or the data entry screen. And then, if you look down to line number eight, we also created definitions for those allowed values, where they were not self-evident. Line number four, database field label. Now, we’re starting to talk about the actual recommendations to the database developers themselves. And you’ll see, that as we go through this, some of our database field labels have two representations. I’ll explain a little bit more detail when we get there, but suffice it to say, we came up with a, what we thought was a simple representation. And then, in certain places, FHIR in particular, the Fast Healthcare Interoperability Resources specified the label, the database field label for that particular data concept. The database field type. Is it a character, is it a date, is it a float integer, et cetera, is also specified, along with business rules, where appropriate. The database allowed values, is a little bit different than the concept of the clinical allowed values, in that the database allowed values, oftentimes, are either coded, or otherwise represented in a database, different than what you would expect to see on a, on a computer screen. The reference ontologies are what we would recommend for the sources for the bindings, for both the concept, as well as the allowed values. You’ll see that some of the allowed values are already included in the value set authority center content, and where that was, where that has been done, those items were listed, as well. FHIR has been a very important and integral part of this particular project; we’ve elected to rely upon FHIR. What we’ve noticed, by the way, is that FHIR, in and of itself, is not completely consistent and so, we’re also viewing this as an opportunity to contribute to further FHIR development. FHIR, though, however, again, has been heavily relied upon as the mechanism to, at least to test compatibility, the approach that we would take to make sure that information that’s captured in a database, could actually be then moved from one system to the next. So with that, let me move on then to talking, actually, to talking
about the actual candidate core CDEs that we’ve developed. Another point, is that we did start with a groundwork, or basis for the work that we’re doing. It is the US Core Data for Interoperability, the USCDI. Here the word core is in capital C. As everybody remembers, this come from the CCDS, the Common Critical Dataset, that antedated the publication of the USCDI. The real question for us, was what out of the USCDI is relevant to registries? And what we ended up with, is this list. You see things that are crossed off. For example, problems, you might say, well, why aren’t problems relevant to registries. Well, they are, but not the way that they’re represented in the USCDI. The typical example then, I would give you, is... Or, typical two examples, are registries are typically setup around that patient already having a problem, like cystic fibrosis, or coronary artery disease, or having received an implantable cardio defibrillator, as opposed to having to extract that patient sample from an electronic health record, based upon a problem list. The second concept, actually, has to do with hypertension, the example that I gave you before, in that if you have hypertension, hypertension has a risk factor for coronary artery disease, is actually defined differently than hypertension as a treatable disorder. So we’ve elected not to include problems, and that’s just kind of one specific example here. As we went through the USCDI, and then took a look at the compilation of the information out of the 38 registries, this is where we ended up. We have eight data, excuse me, data elements, data concepts that are more or less as is, from the USCDI. We’ve already talked, by the way, about four of them, but we’ll spend a little bit more time talking about them today. But in our previous two meetings, both the webinar, as well as the face-to-face meeting, that we had at the Pew Charitable Trust last month, we’ve already mentioned sex, race, ethnicity, and smoking status. And then to our seven adjusted concepts, where there are modifications that are based upon our work, based upon what’s been represented in the registry forms, et cetera, and you can see the list there. They include three items at the bottom, that are not in the USCDI, which we believe are important. They are at least represented in the registries at some level, and so, we spent some time talking, or excuse me, developing those concepts, as well. So with that, I would ask you also to bear with me. We’re gonna walk you through most of these items here. It turns out, that to represent these 15 concepts, requires about 60 different data elements. We’re not gonna go through all 60 here, we don't have enough time to do that. So what I'm gonna do, is be, again, kind of selective, pick off some of the easy ones. Also talk about some of the more difficult ones, and, hopefully, set it up so that you can then review the spreadsheet, and then return final comments to me. Now, I’m not, I'm purposefully not presenting to you the spreadsheet, because you can’t read the spreadsheet on a screen. So what we've done, is we've extracted all the information from the spreadsheet, and then made them a series of slides here. I think, for those of you who worked in the CDE arena before, you'll kind of have an idea about what I’m actually talking about. But if you can think of these 12 items here, these are the pieces of metadata, as the rows across a spreadsheet. And then the concept, like for example, patient name, as an individual row of the spreadsheet, then you can start to think about how this would actually look on a spreadsheet format. But again, just to save you all the eye chart based approach, trying to figure this out, we’re gonna go through in fairly rapid-fire, the work that we’ve done, one data element at a time with the slides, trying to capture the information that we’ve created. And so, for patient first name, so the first thing is that there is a concept, obviously, patient name. We separated that into three concepts. One is the patient first name, the second one is the patient’s last name, or family name, and then the suffix, if appropriate. Now, you could say, well, what about the, the prefixes, what about everything else, middle name, et cetera? Remember, this is a minimum set of concepts, not everything, but what we think is the best starting point. If a registry, for example, doesn't even capture a name, has a different way of managing that patient identifier, then you
wouldn't need to think about this. So we're not asking you as a registry, to develop all of the contents in your registry. We're asking you, if you have that particular concept represented in your registry, this is how we would suggest that you, that you include it. Looking at patient first name, the clinical concept label. This is what we would recommend, that would be on a case report form, or on a computer screen.

We would recommend the label, given name. Now, you'll see in brackets, that there is also first name. We recognize that, and you'll see this for most data elements, that there is some variability. What we tried to do, was pick the ones that both A, appeared most frequently, as well as B, were, I guess I would call 'em semantically the most accurate, but also, the most understandable. So again, if you capture individual patient names in your registry, we would suggest that the term given name, which is much more a universal application of the concept of first name, than just the words first name, if you think especially universally across the world, that's the label that we would suggest. Going down to line number four, the database field label, you'll see that there are two representations here. The first one, is the simple representation, that's the, if you will, the way we would suggest that a database represent that particular field in the database, GIVEN_NAME. You'll notice that there's a convention that we're following, it's all caps. There are some databases, that do not handle upper case and lower case all that well, and because there are no databases that really behave well when you use spaces, there's an underscore there. The second of the database field labels, the Patient.name.given, is taken directly from FHIR. One of the things that, excuse me, one of the things to suggest, is that if you have a database that can handle both the period, as well as upper and lower case, that that actually be what you do, because then you'll be FHIR, you'll be FHIR ready already, because that's how FHIR expects the data to be transformed, as it moves it from one system to another. Line number five, I'm open for suggestions. We didn't know how many characters to allow, we talked about 255, we talked about a more limited number. You might wonder where the number 140 came from, that's the original limit from Twitter, for tweets. So we thought that would be kind of cute, and probably encompass more than needed. There is a little bit of humor that we built into this particular project. Again, I'd be happy to solicit any opinion. All right, we're gonna go a little faster now. I think you get the basic framework here. We did try to include at the bottom, when you see number 11, the FHIR references, and if there's any additional things that you need to know, as well as the binding, the primary binding in this particular case, the LOINC, 45392-8, is termed the first name. The patient's last name, or family name. The suggested label is family name. The alternate is last name. Again, you see in line number four, FAMILY_NAME, and then Patient.name.family. Why 70 characters? Well, we don't think last names are as long as the potential, at least, for multiple first names, so that's half of a tweet. That again, is open for further discussion, or suggestion. But, by the way, if we arbitrarily came up with something like 70 and 140, it's because we couldn't find a standard out there. We were trying to be all encompassing, but, it does represent a fair amount of digging, trying to find the information. The only other piece of name, that we thought was relevant, to make sure that we could identify individuals, was a suffix. And so, therefore, we included suffix. The FHIR, in line number four, it does include the suffix concept, Patient.name.suffix. And then, the rest of the concepts there, are as you see. Excuse me for a minute. All right. Next item is sex, we've already chatted about Sex, in several of our meetings. The concept of sex is specific to the biological sex of the patient. We felt that that was more important, or more relevant, because that's what would be likely predictive of clinical events, as opposed to the social construct of gender. Again, the concepts are there. What I'm gonna start doing now, is just chatting about each one of these concepts, but not, not reading a whole lot of it, unless there's a specific point that I need to point out, in terms of the content that you see. Race, I would ask you to look at line number four, again. You might say, why is there caps
RACE, and a little race. Again, this is just to point out, that the simple name, the simple database name that we’re suggesting, follow this all caps format. There is a FHIR representation for race. Interestingly, it’s just race, it isn’t patient.race, or anything else that would have a longer representation. You see the allowed values, there are no definitions of those allowed values, mainly because, they are self-evident, and to, to include definitions of something that’s self evident, just doesn’t make a whole lot of sense. You might think, to yourself, for those of you who’ve worked in this area, where is the VSAC representation of them, you know, who is CDC PHIN? The CDC, obviously, is the CDC in Atlanta. The PHIN is the... well, now I’ve forgotten what PHIN is. Population Health Information Network. That’s where the value set comes from, originally. So rather than referencing the VSAC, which references the CDC PHIN VADS, we said, well, we oughta just represent CDC PHIN VADS directly. Ethnicity follows the same concept. And again, I’m not gonna belabor the point. The Office of Management and Budget is the group that is identified. Hispanic or Latino is the ethnicity that needs to be specifically identified. The only things that I would comment about both race and ethnicity, race, for example, has in the detailed list, over, I think it’s over 800 different races. Obviously, we’re not representing all 800. And then the second thing about race, is that the Office of Management and Budget has these as discrete concepts, as opposed to ones where you could have combinations. That’s something that’s still out there, in terms of how to capture the information. But at least we’ve gone through, and provided the framework for that discussion. Next concept is date of birth. I’m not sure why we have something in red here. We made this slide very quickly. Actually, not very quickly, we had to, I thought we went through these over and over again, but at any rate, line number 10, I think, is something that we just need to clean up. It is cleaned up in the spreadsheet, so when you get the spreadsheet, you'll see that. But date of birth is, again, a very simple concept. It has, in line number four, two different representations, BIRTH_DATE, and then, Patient.birthDate, and it follows a date format. Smoking status. Smoking status, we’ve already chatted about. I’m personally not particularly happy with how the CDC has defined this. The concepts are defined, current heavy tobacco smoker. We have purposely not included all those allowed value definitions, because it takes up too much space, but suffice it to say, it will be in the spreadsheet for you to review. Device identifiers. This is about the unique device identifier. The parenthetical comment I’ll make here, is that what is needed for registries, is not the entire UDI. The entire UDI, for those of you who are not aware, the UDI itself, is a, is a unique device identifier, for every device manufactured, medical device manufactured. And the class threes, and the implantable ones, there is regulation, requiring the inclusion of the UDI in the patient’s electronic health record. The UDI itself includes two components. It includes a device identifier, and a production identifier. Production identifiers can be thought of as the serial number. Oftentimes, other pieces of information that could be identified with an individual patient, as opposed to a class of devices, like all of the 2.5 millimeter by 22 millimeter stents. So the DI component would be at the level of a manufacturer, the Device Identifier would be at the level of the manufacturer, and the class of device, like again, 2.5 by 22 millimeter stent. The PI, the Production Identifier, could potentially go down all the way to a serialization of a given device, such that it could be used to identify an individual. So if you think about it from a registry standpoint, what’s really, at least what we believe is important, is the DI, the Device Identifier component of the Unique Device Identifier. And so the concept label is the UDI DI, the Unique Device Identifier, DI, the Device Identifier itself, not the entirety of the UDI. HL7 has been working on the movement of this piece of information from, or among systems, and there’s already a representation at the database field level. And that’s item number four, Device.udi.deviceidentifier, and you can see that there. All right. Onto some other things, that I think you’ll be very comfortable with. One is... The next set, is the things that
are related to the vital signs. So this is height. I don't think it takes rocket science to understand the concept of height. I'll just sit here for a moment, and let you guys read through this. You'll notice that there are blank things. We debated as to whether or not we wanted to include in the database field, if you will, business rules, a minimum and a maximum, and elected not to, to be honest with you. Part of it is, is that there's different ways to capture information, like centimeters and inches, which we'll get to in just a moment. And so, the ranges would be dependent upon what your unit of measure is. But part of it, was that if you think about it, that's something that people can develop, but this is not that often, mistakes aren't made that often, in this particular arena. When you're talking about height, I mentioned before, that you do need to have a knowledge of the unit of measure that you're using. And so, this is where we get into units of measure. You'll see that for height, and then for the next one, weight, we are using the UCUM standards, which are referenced by FHIR for the units of measure for height. If you were paying really close attention, if you look at line number seven, where it says database allowed values, it says cm, which is, obviously, stands for centimeters, in, which is inches. And then there's kind of a funny thing, called in_i. Well, that's the FHIR standard, it's inches international. I'm trying to think of where inches might be measured differently, or mean a different thing anywhere else. Suffice it to say, that's in FHIR, so we included it, but it's a bracketed recommendation, which means it's an optional thing, as opposed to something that's required. Same kind of rules with weight. There's a definition for weight, BODYWEIGHT, and BodyWeight, in line number four. The LOINC ontology concept is listed there, and then the weight units are listed here. On line number seven, the database allowed values, kilograms, grams, pounds, and then pounds and av. I am blocking what it stands for, but it basically means it's US and England. That's the approach to weight units of measure. The next vital sign is heart rate. Heart rate is shown here. The optional label would be pulse. We debated whether it would be pulse rate, we actually looked it up in several different record, reference sources, and the word pulse actually is a rate. We specified that the number of heartbeats per minute would be the way that you would capture the heart rate. After all, that's how it's done anyway. And so, this then raises the question, do you need the units of measure? And we've elected not to, since heart rate is always in the number of heartbeats per minute, as opposed to the weight of a person, and the height of a person, where you need to know what the units of measure are, we did not include units of measure for heart rate. Same rules, then, for diastolic blood pressure; diastolic blood pressure is always measured in millimeters of mercury. Now interestingly enough, FHIR does not have a way to uniquely identify blood pressure diastolic, or diastolic blood pressure, as a single individual data element. It's a combination of data elements that are required, rather than a single data element, so now you see a convention where we're not using anything, you know, on line number four, where it says database field level, after that, it's just BP_DIASTOLIC. There's no comma, and no FHIR-based representation, because there is not a... There is not a single FHIR-based representation. We're gonna try to clear up this noise here. All right. For diastolic blood pressure, and then for systolic blood pressure, we followed that convention, that where there is not a single database element that would be appropriate for database developers. Then we have only the database field label, that's the simple version that we have recommended. Next concept is death date. The date of death is something that we thought was really important, especially as we're trying to think about what registries are capturing, or we're assessing, or we're evaluating, what registries frequently capture the, the occurrence of death, and especially, the date of death is something that we saw fairly frequently. Interestingly enough, again, in FHIR, it is under development. So if you look at what lab... Excuse me, if you look at line number four, you see that the simple representation is DEATH_DATE, and then the FHIR representation is Patient.deceased, but there's an x there, because
there is still development work that's ongoing within the FHIR organization, to specify how a date of death would be represented. Although, you can find the information about death in FHIR, it is still not completely vetted, from a standpoint of how to represent a date of death. A little bit different than the concept of systolic blood pressure, where you can represent it, but it isn't a single element, here there is no, at least right now, recommended way to recom-- to document in a database setting, the date of death. Next concept, is related to the care team. But we didn't think that the entirety of the care team really was that relevant to registries. As registries are constructed, the procedure-based registries frequently inquire about the operator, who the physician was that actually performed the procedure. And because of that, we felt that it was important to capture that piece of information. That information, oftentimes, is used to develop experience curves, that is how much experience you need to become, if you will proficient, or facile at a procedure. It can also be used from a quality standpoint, for determining the appropriateness of procedures, etc., etc. So we did need to have a concept to capture information about the proceduralist, so that's where the NPI comes into place, the National Provider Identifier. The reference at the bottom, the NPI Registry at CMS, is where the actual values for the proceduralist would be captured. We struggled a bit with this one, and this is, the concept is very simple. The concept is, who's your doctor? Who's the one who's making the recommendations for, or that you undergo a diagnostic test, that you get a chest x-ray, or an EKG, or that you have a beta blocker written for you? Using a simple set of words, to say who's your doctor, was a little tricky. So where we've ended up is just saying the clinician NPI, with the definition being the National Provider Identifier of the clinician to whom management of a patient is attributed. That is, who's making the management recommendations for that given patient. So it's the attending physician, it's the NP or PA who you're seeing in the clinic, etc., etc. We did think that that was important, relevant to decision making, relevant, potentially, to outcomes, relevant, obviously, to compliance with guidelines, evidence-based medicine, etc. If anybody has a better frame, a better term to use for this particular framing, I would invite you to send along that information. Procedures, procedures become quite important. We ended up arguing that the place where there is the greatest consistency for procedures, is with CPT coding. There are other coding schemas, in fact, there's... The problem with procedures, is that there's a half a dozen coding schemas, including various forms of ICD. The SNOMED folks have procedure coding terminologies, etc. Since billing is universal, this is where we've ended up with, in terms of recommending the procedure code, that the procedure code be the billing code, for the specific procedures that are performed. This is oftentimes less important, interestingly, for procedure-based, for procedure-based registries, mainly because, you already know what procedure has been performed. But where it comes up frequently, especially in the procedure-based registries, is when other things are done. And that's where we're thinking that having procedure codes would be important. Procedure date is not, is something that's aligned with procedure code, but the procedure date really is about the primary procedure, not the additional procedures that are done, and that's this concept here. The last two concepts that I want to just chat about briefly, and then I'm gonna turn it over to Tom Windle, for him to talk about the medications model. The first one is, illicit drug use. We did see illicit drug use appear in several of the registries. We used a combination of Wikipedia, and our common sense, to come up with the clinical definition, because we really couldn't find one singular definition that was out there. The summary of the use of illicit drugs, comes from a 1975 paper. This is something that's a NIDA based paper, the National Institutes for Drug Abuse paper, or Drug Abuse helped with the authoring of this particular framing. But it is old, it's from 1975. Now, NIDA, interestingly, has a fairly large ontology, that they use for classifying the drugs that are used. We're going to include that in the, in the final
report, in the final representation. We were trying to get that ready for today, but we didn't quite have it ready. But, suffice it to say, that will also appear. And then, finally, alcohol use. Alcohol use did appear in a couple of registries. It's not particularly sophisticated, and in fact, again, we did not find much in the way of reference content, aside from NIDA, the National Institute for Drug Abuse, in terms of alcohol use. We were thinking that they would say number of ounces of alcohol, or something along those lines. But, actually, even the NIDA surveys just ask the question: Did you have anything? Did you have any beverage with alcohol? And that's what they call a drink. So we left it at less than one drink per week, two to seven drinks per week, greater than eight drinks per week, none, or unknown. Okay. So, with that, that's the quick run-through of the data concepts that we've developed, with the exception of two big groups. One is the medications model, and the second one is the labs. We are only gonna take a look at the medications model, as a, a way to think about content where there's multiple different dimensions that need to be captured, to, to express the, or attribute the use of a given agent. In this particular case, medications or to get a result back, in the case of laboratories. We're only gonna look at medications. We will have labs represented shortly, although, I don't think we'll have it in time for your homework assignment. But, suffice it to say, that's the last kind of big piece that we're working on. With that, let me hand it over to Tom Windle. He has about a few more minutes worth of stuff to talk about, regarding medications. And then, we'll finish up with your, with what we're asking of you, the comments and next steps. So, Tom, if I could hand it over to you, for some comments.

- [Tom] Good afternoon. So while going through all of the registries, we discovered there were four general types of questions that were being asked. Some of you have seen this before. It's provide a list of medications, which is almost, is kind of an unstructured way of just asking for what were they on? Specific, domain-pertinent medications. If they ask if they are on a TOBI Podhaler, for instance, something very specific. Medications in a class of drugs were the most common, that was, is this patient on an anticoagulation drug, for instance. And then, the final category, was medications administered related to a procedure, either pre, during, or after, for that. And there were six major, or there were 16 major units, questions that could be asked in registries. But there were six, generally, that were found over and over again. And we built out a little bit, a model for that, the medication name, dose, units, source, code, and class. So when I go to the medication name, we built this out in FHIR, and it's to identify the name of the medication being administered. This links the resource that represents the medication, which may be the details of the medication. And so, essentially, what they're saying is, just overall, what's the name, not related to packaging, dosing, or quantity, anything like that. Interestingly enough, it is a subset of medication code in FHIR. So it's Medication.code.text. And once again, similar to others, this is a, they are soliciting feedback in FHIR, for their rules, for the pharmacy. So if we have anything better, we can submit it to them. If we get any ideas from this. So, yeah, the medication name was pretty straightforward. I did a string of 128, that was the longest string I found. I'm thinking about it right now, that's probably insufficient, because, I didn't include medication combinations. Yeah. For right now, it's 128 string. The next one's a little bit more complicated, the medication code. Now we had this as a code that is specific to a medication, and it's a 12, or 11 digit code, in RxNorm, that we can roll-up to NDC codes, depending on it. So the RxNorm code, has a lot of very specific, can handle a lot of very specific information, but it is not pharmacy-level medications. So if we wanted to know the, the pharma, the distribution of a medication, including a manufacturer, and packaging, we'd want to use NDC. So this only applies to clinical prescription. I didn't see any questions specifically in any of the registries,
regarding distributed medications. But, we do have that in mind, while we're building out our rules. Generally, we are just for Rx class. And then, medication class, medication class is the next one. We got a lot of very generalized classes, that were asked about. Specifically anticoagulation, antiplatelet, I mean, you can see the list here, it's a lot of different classes. But very rarely were they codified into a Rx class format, only four had any sort of stated codification behind it. A lot of 'em made their own list, which is nice. But we want to help you guys get out of the, the business of drug librarian, and more into your specialties. And so, that's part of the reason we're doing this, to try to get it so that you have a lot more information that's available to you. Have a lot more resources that allows you to ask questions without having to worry too much about how you self-define it, that the definition will be there in the first place. And I think that's what we're really on a good path to doing, in our recommendation lists.

- [Dr. Tcheng] Great, so thank you, very much, Tom. We are basically finished, or just about finished. I do want to have an opportunity to at least address a few of the questions that have popped up. Before we do that, let me just have Becky finish up with the presentation here, with our next steps. She's our fearless leader, and she's been the brain child behind all of this work, so I want to give her due credit. But, if you'll just finish up, and then, we'll address some of the questions.

- [Rebecca] So, in terms of next steps, the main thing that we want you to know is to expect an email later this week. It'll have a link to a survey that we're gonna use to collect feedback. We'll have a few specific questions for you to address, as well as a place for general comments. We'd like to get that feedback in by the 15th of October so that we can incorporate it into the final work products and the final progress report. The final progress report will be coming out later this year, in December, and then the manuscript, we will be selecting a target journal and a date of publication, working on that after we get the progress report done. The main takeaway is that there will be a survey coming to you this week, along with the spreadsheet, and we would appreciate your feedback. If for some reason you want to get other people from your organization to give feedback that would be welcome and you can forward the link and the materials to them as well. So, thank you, Dr. Tcheng.

- [Dr. Tcheng] So, with that, I'm gonna walk around here a little bit because I've got to get to a different computer. But, we'll take a look at some of the questions. All right. I'm not gonna identify the individuals who are asking the questions. I do want to just--

- They are identified.

- [Dr. Tcheng] Oh, they're already identified, all right. Well, never mind, so, everybody's already identified, so we can just go ahead. From Kevin Baskin, at VANGUARD: Would still like to see a lay definition appropriate for use by, and understandable to affected persons, such as patients, family, caregivers, support organizations. Is there a plan for this? So, Kevin, I'll turn that around to you, and I'll say, we tried, or I tried, and I think we collectively tried to use relatively simple, straightforward
definitions. If there are things that are not clear, especially to a lay individual, please, point those out, and, yes, we can work on these, especially the clinical definitions. The clinical definitions are not derived from a specific source. They are, basically, original to us. That's where we're being a bit creative. I think the answer to that question, is yes.

So, from Al Taylor: Some registries are not setup around specific problems. Examples include primary care registries, like PRIME. That's an excellent point. Now, I think what we are trying to do is make sure that there's a base vocabulary, a base set of data elements, that it really doesn't matter what your registry is focused on. That we would anticipate that you could develop the concepts that are part of this core data element set that per the recommendations that we see here. Even if you're registry is not focused on a specific issue, or specific problem, that doesn't mean that you couldn't incorporate the things that we have here. Stated another way, if you think about the way we set this project up, what we were looking for was commonalities across as many registries as possible. And so, that's the criteria that we used to pick the CDEs, not the other way around, but, your point's well taken.

From Dr. Huser: The FHIR Reference is too broad. We actually are including not just global FHIR, but, the actual place in FHIR where you will find the content that is needed to assure that the database elements actually are interoperable. We purposely tried to make sure we identified not just the profile but also the resource. The slides are only showing you a portion of that work. Oftentimes, we go down to the level of the individual page. There is a very specific tab in most of the FHIR resources that has additional details and that's where I would ask folks to go. We'll have that in the instructions but I think we've been pretty good about the URLs going to the actual resources, not just the profiles.

Okay, from Al Taylor: Biological sex is not equivalent to sex assigned at birth. Perhaps modify the definition to align with other birth sex definitions like what is in HL7 PHIN VAD. We've struggled with this, and again, Al, I will ask you to think about this. You're absolutely right, we did, we avoided the word genetic sex, which most people have not, well, I can't say that today. You know, when you can send your mouth swab over to 23andMe, and then get it back, maybe a lot of people do know what their genetic sex is. But, if you look at the concept of sex, and I'll reference Wikipedia right now, it's the concept of the sex that's assigned at birth, not a genetic sex. We'll take this under advisement.

So what's our time here? I probably should try to answer these quickly, here. All right. So many patients now are mixed-race, we don't identify with one or the other. Yeah, I agree with that completely. One of the things that we are contemplating, is... You know, I mentioned this before, that there is 800+ races out there, so that's one aspect of it, that's not covered in this. The second aspect is how to get away from the OMB's construct of everybody being a single race, to combinations of race, or, you know, not being even able to identify. We, I think, will ultimately make the recommendation, that this not be a exclusionary, or unique, you have to choose one, but that you actually have the optionality of choosing multiple.

All right. PHIN VAD, so, thank you, Wendy Blumenthal. Let's see, what information value sets in... Yes, that's correct. The value sets that are in the PHIN VADs are not necessarily defined by the CDC. I apologize, if I misstated the fact there, but it's another standards development organization.

All right, Dr. Huser: Consider a metadata field example value.
- Adding a metadata field, and giving a value.

- Oh, oh, adding a metadata field, okay.

- Giving an example.

- Okay. Yeah, that's a great, that's a great idea. Yeah, that won't take too much time, we can, I think that... That suggestion is something that we can develop pretty easily.

Try applying height to an infant. So height is not defined as having to stand up. Height is defined as from the bottom of your feet, to the top of your head, however you would define that. It's a mechanical measure, as opposed to a one where it's required that you administer sedation to the infant, to get him to lay still. But, any rate, the... We'll leave the technical administration of the determination of height, to those who are actually trying to determine the height.

Okay. Don't specify expected units.

- [Rebecca] Adding a business rule.

- [Dr. Tcheng] Oh, yeah, adding a business rule, okay. That make sense, all right.

Let's see, from Kevin Baskin: Wouldn't some of these concepts be incorporated into the relevant clinical data models? You would think that they would be. What we're planning on doing is taking the work here, and pushing it, putting it through the Clinical Information Interoperability Council, Health Services Platform Consortium modeling approach, so that it actually is modeled in CIMI, the Clinical Information's Modeling Initiative. That includes the work that's necessary to make sure that, that, again, that the complete approach to representing these concepts has been, has been followed. This includes building additional concepts in LOINC, if they're necessary, or any other of the standards development organizations. So, yeah, what we're talking about here really is the framework for doing that kind of work. The CIMI modeling process really does require good strong metadata to start off with as the starting point. And so, that's how I believe we're gonna be contributing to data models that then be incorporated, formal data models, that can then be incorporated into electronic health record systems and solutions.

All right, Wendy Blumenthal says: The smoking status value set in PHIN VADS is the one defined in Meaningful Use. Yes, that's correct. It's terrible, that's the problem. I know it's, and that's why we've developed it, it's, basically, explicitly as is in Meaningful Use, as is in represented in PHIN VADs. The question for the group, though, is, is that what we really ought to be using, at the clinical registry level? And it's more of a open question to solicit opinion, as opposed to a commentary about what's been done in the past. And I'm probably speaking as a cardiologist more than anything else because we've
tried to use the meaningful use recommendation, and have found it relatively inadequate for our purposes. And so, I just throw that question out there.

Okay. Why not submit this to LOINC? LOINC is already a big part of this. I would say that, ultimately, the formalisms for informatics modeling actually don’t fall at the development of standards development organizations, where they create bindings to the concepts, etc. What needs to be done next is go from where we are, which is the bindings, for example, especially with LOINC, also with other SDOs, and then build these things out so that electronic health record vendors, database developers, et cetera, can actually incorporate these concepts and physically build them into the databases. That’s the framing of this project, right from the very beginning, is how do we go from an SDO binding based approach, to something that can be physically built into databases?

All right? Was there any consideration to following patients across registries, or the affects comorbidities on a disease condition? Two different questions. Across registries you need patient identifiers, so, yes, I think that this does set it up, although, I would suggest that the political, legal, ethical, and moral issues are much greater than the technical issues related to following patients across registries. This project definitely is not attacking any of those, of those issues. And then, the affects of comorbidities on a disease condition? We were actually looking for that across registries. Now remember the construct, the, originally the grant proposal was to find things that were similar across registries. And because, for the most part, the comorbidities are different from disease to disease, from procedure to procedure, et cetera, we actually did not find much in the way of commonalities. The only commonalities, as you might anticipate, are for example, the Society of Thoracic Surgeons, and the American College of Cardiology, have a shared list of comorbidities. But that’s because we’re all dealing with coronary artery disease. The answer to the question is that would be nice, but that’s not what we found. So we’re gonna stick with what we found as a starting point, the basis for the work that we’ve done. So with that, I think we’re gonna call it to a close.

We’re a couple of minutes over, I apologize about that, but I did try to walk through as many questions as quickly as possible. I do appreciate, or we do appreciate, the whole team appreciates everybody’s participation and support for this project. We have the opportunity to really start this interoperability engine, going full steam. And you, obviously, have to start somewhere. You might say, well, this is pretty simple stuff, but this is how it’s done. You start some, you start at a place, and then, hopefully, continue the development using this as a paradigm. So again, thank you, very much, for your attention and your time. Please be on the lookout for the email with the preliminary draft of the spreadsheet content. It will not be the final draft because I’m sure folks will find issues, that we will continue to develop it. We’re gonna take the extra time and do the best job possible. Our email addresses are on this, on the screen now, and our project website is at the bottom. Again, thank you very much for your participation, and have a good rest of the day.