Pew-DCRI Healthcare Interoperability Stakeholder Meeting
21AUG2018
Morning Session

- [Anqi] Hi, everyone. I think we're gonna go ahead and get started. Thank you all for coming at the Pew. My name is Anqi Lu. I work on the Health IT Team here at Pew, and we are really excited and lucky to be partnering with the Duke Clinical Research Institute on this project. So, before we get started, just a few logistical things. I'm sure most of you have found by now, but if not, the restrooms are out of this room, if you turn right, the gentlemen's room is before the elevator, and the ladies' room is right after the elevator. We'll have some intermittent breaks throughout the day and so, feel free to get up and move about. Lunch and all of that will be served outside in the big place where the breakfast is and feel free to eat and drink in here as well. A little bit about Pew for those of you who may not be as familiar. We are a large nonprofit research and advocacy organization and we have a bunch of projects and a whole suite of different areas, some including environmental work across the globe. We also have a lot of state, mental health, consumer finance, pensions, criminal justice reform, all of that. And we also have a very large healthcare portfolio, which I'm a part of. So, here from the Pew Health IT Team, we also have Ben Moscovitch who directs our work on Health IT. And we have Josh Rising who leads a bunch of our healthcare projects, including the substance use and treatment program. What is it? I can't even remember what the acronym stands for.

- [Male Participant] Prevention and Treatment.

- [Anqi] Prevention and Treatment. Yes, thank you. The prevention comes first. And also, our Health IT work as well. So, we have projects in our healthcare work and a bunch of those areas, also, and our group is into safety and so on. So, that's a little bit about Pew. I think with that, I'm going to turn it over to Jimmy to talk a little bit more about this project and so welcome Jimmy.

- [Dr Tcheng] So, while Anqi's working on trying to figure out the tech AV stuff here. I did want to again, acknowledge the sponsorship of Pew in making this all happen. Anqi, Ben, and Seth are the brains. A trust fund from Pew that has enabled the work that we're going to be sharing with you today. I also want to publicly acknowledge my brain trust, the folks that I've relied upon to try to help move this forward, John and Tom Windle from the University of Nebraska. Where are you John? Over there. Joe Drozda from Mercy Frontier. Julia Skapik from Cognitive Medical. Thank you. Seth Blumenthal. I saw Seth, he's with AMA. Marjorie Rallins and Crystal Price from the PCPI. And obviously my good team, too who you will get fairly acquainted with as we go through the morning. That needs to go there. Is it a lot of stuff? It appears so. Okay. Oh, that's good. Alright. All right. So, I'm just gonna use a few slides here to try to set the stage for what it is that we're going to be attempting to accomplish here. This is a Twitter feed from a Twitterverse, a colleague of mine found and I think it's quite apropos that the most unbelievable aspect of the Star Trek universe is that every ship they meet has compatible video-conferencing facilities. So, if you think about it, that really is the dream. We can do this. All right? Now, this is obviously science fiction and now, the question is, how do we change this into becoming, if you
will, science fact? It isn't for a lack of trying, so a little bit of history. The first reference that I could find to an ontology in the setting of medicine dates back to not the 19th, or not the 20th Century, but actually the 19th Century. In 1965, SNOP was published by CAP, the College of American Pathology. In 1966, the AMA published the first edition of the CPT codes. HL7 was, I guess you could call it, launched, depends upon who you believe when... Ed Hammond's at Duke, so I'm the Duke PI, you'd have to listen to Ed and he always claims credit for being the person who launched HL7, that's why I put his name in parentheses, because I know it was a team effort. LOINC, Clem McDonald. 2011, Grahame Grieve brought forward the ideas that have now become FHIR. But, in reality, I would suggest to you that although all this work on the development of ontologies has been very, very, a lot of resources, a lot of energy have been paid towards it we still are a far cry from being able to capture information on the clinical side, and then especially get it over to the registry. So, I'm gonna ask you guys a couple of questions. So, the first one is, how many biomedical ontologies have actually been developed as of yesterday, when I last checked? And this is reference data from the bio-ontology portal, that's where I'm getting this. Anybody have an idea? So, let's see a show of hands. How many think it's about a hundred different ontologies? 250? 500? 750? Y'all are doing the medical student thing, which is the answer in the middle is usually about the right answer. Okay, 1000? Actually the answer is 750, more than 750 ontologies, think about that. Different ways to classify what we do in medicine. And then, how many terms have been developed or modeled as CDEs, data elements. Now, there's relative degrees of formalism and some are much more formal than others, but in these ontologies, how many would you say have been actually modeled? How many terms? A million? Again, a show of hands. Five million? 10 million? 20 million? Any 40 million takers? All right. Bob, I think you get the prize here. It's 40 million. All right? So, pretty impressive. So, this is an interesting site. I don't know if this will work here or not, if I click on this, but I would ask you guys to bookmark this. You can find every ontology that there is known. This is a repository managed by Bioontology.org and the interesting thing is when, so there's 726 ontologies that are listed in bioportal.bioontology.org. There are 39 million indexed records, you look at a record, that basically is an entry, okay? Not accounting for overlap, cetera. But, if you go in here and then say, "Let's find an ontology, like cardiology and you ask, so C A R D I, or C A R, yeah, C A R D I. How many cardiology ontologies are there? There's actually four of them that are listed there. One of them is an ontology, Cardiovascular Drug Adverse Events. And, Ganesan. Ganesan's from, Ganesan Srinivasan from the ACC, the American College of Cardiology, so I'm going to pick on him for a moment. I know you guys can't see this, but none of the societies that are representing the house of cardiology, the ACC, et cetera, are in there, okay? So, we have our own data dictionaries and registries, et cetera. And, this isn't just true for ACC, I can pick on any society, just about any society here and you'll see that even though there is a formal way to manage the Bioontologies, they're not being represented, at least, in BioPortal, all right? Just trying to get back out here, close this window. So, where was I? So, Terri Reed and myself at the FDA have been trying to put together, at least from a clinical perspective, a roadmap of what all this looks like from the standpoint of their action and we've come up with this concept that right in the middle, there used to be a playbook. The playbook has the general, or core common data elements, which is really going to be the focus of what we're talking about today, that should be shared across, really, every health information technology system. And then, depending upon if you're a cardiologist, or ontologist or physical therapist, or whatever it might be, there are going to be domain-specific common data elements. The UDI is separated out, unique device identifier, because of its central role at FDA. From an outcomes standpoint, the Agency for Healthcare Research and Quality spent a lot of time and effort and money trying to figure out outcomes, so we've actually kind of
partitioned them. So, what I want you to see is that there's a lot of folks here circling around these concepts and they have various roles and aspects in terms of thinking about this environment. Let me see if I can get it. So, in the inimitable words of Laurence Peter. Everybody knows who Laurence Peter was, right? The author of the Peter Principle, okay? This is his less famous saying and that's, "Some problems are so complex "that you have to be highly intelligent "and well informed just to be undecided about them." And that's kind of what we are talking about here, is that how do we fit into an ecosystem of things that we need to accomplish in order to improve interoperability? So, the problem statement, to me, is pretty straightforward. The first one is interoperability of what? So, I know we have representation here from ONC. I would remind everybody that if you think about what the larger mission of healthcare really is to move information around, but ONC is largely focused on the interoperability of PDFs, but that's not about the data. So, how do we get to interoperability of actual data, and interoperability of clinical data, so you can move it from one system to the next. Not just PDFs for human eyeball consumption, but actually moving the data around. What will it take to make interoperability a reality? And I would suggest to you that 120 odd years of building ontologies is probably a good start, but it's not the answer because if it were that easy, we would be there already. And then, the reason to convene everybody here, not just our registry partners, but the federal government, the federal representation here is, what can the registry community do to accelerate interoperability? So, that's the problem statement. What I'd like to try to accomplish today. There is an element of audacity here. Audacity includes the willingness to take bold risks and despite the fact that you might think this is pretty straightforward stuff. Actually, this is revisionist history here. We're actually asking for a transformation about the way, at least the registry community, approaches managing day-to-day elements. The second part of audacity is it's rude and disrespectful, so I've already abused the ONC. Apologies, okay? And by the end of the day, I'm sure I will have affronted a number of folks here. But, the real question is, how can we move away from this kind of me, me, me type of environment? and say, "Let's do this all together "and accomplish something that nobody has yet accomplished." So, what are the meeting objectives? The meeting objectives are the following, to develop a deep understanding of what it is that we're trying to accomplish through this project. Give you a glimpse, a snapshot of the work that we've done to date. We do want to contribute, we do want to have you contribute to the discussions about the strategies that are going to be required in order to move us from where we are today to where we really should be and to understand, to have you understand your role in the next steps. And, that is to understand that there is a work product that we're going to be pushing out to everybody here who's participating and asking your response. That is, do you agree with this? And then, for those who are registry representatives, position yourselves in a way that can actually bring forward the adoption of the core CDEs that are the project here to your respective registry. So, with that, I'm gonna stop and turn it over to Becky. But, before I do that, any specific questions at this point, at this juncture? We do want to have this as open of a meeting as possible, so if things come up as we move forward, then please raise your hand and speak up. Seth?

- [Seth] Hi, Seth Blumenthal from AMA, just want to, I just think of this concept from real estate. You know, sometimes you build a building and you already have a webpage signed up to enter, sometimes you build the building on spec. So in this case, we're building and offering for use soon, these 20 or 30
common clinical data elements on spec. In a sense that we don't necessarily have a use for them yet. So, something for all of us to think about today is, especially for those of you that are involved in registries and other development data systems, is if you have these data sets, if you have the ability to download them and put them in your system, implement them in your system today, this month, what might be some of the benefits of doing that? If you were going to study that, what might be some measures that you would put forth to that effort?

- [Jimmy] Okay. Thank you.

- [Becky] So, good morning and welcome. We're all so glad to see all of you here today. Thank you for being here. So, I wanted to go over the goals and objectives for the project, just, quickly to set the stage for our discussion. We wanted to find out one of the primary goals of the project is to try to describe the landscape of how registries are implementing standards today. We want to be able to describe for you what some of the different standards, initiatives that are going on and help everybody figure out how to pull all the pieces together. And we wanted to be able to create an implementation guide for the data elements that are recommended, so that we can advance interoperability. We're gonna accomplish this through three aims. First, with a case report form review and abstraction of the data elements that you're all already collecting. Second, we're gonna try to look at how you're collecting those information compared to the national data standards that are available. And we're also gonna try to come out of that with a synthesis of best-practice recommendations for everyone. We've got over 38 registries participating in this initiative and we appreciate your efforts and contributions to do so. Collectively, everybody's efforts will be greater because of your participation. And the deliverables from the project are obviously the data element recommendations and the environmental scan to show you all the different things that are going on in the standards space. The particular concepts of interest that we're focusing on are the ones that are common across all registries, or that most registries collect. And so, you can see the list here. I won't belabor the point, you'll hear more about these as we move forward. And we're also gonna be able to incorporate the metadata that would be necessary to implement these data elements in registries and HIT systems. And so, these are some of the core metadata pieces that we're gonna be focusing on, but open to feedback if others are needed. So, the recommendations and the environmental scan will be coming forward and you'll hear more about those today. At this time, I'll turn it over to Julia unless there are questions for me before we move ahead. Julia?

- [Julia] Okay. So, I didn't collude with Jimmy on what I'm going to talk about, but his slide sort of perfectly set me up for what I'm gonna say here. So, my goal was to talk about, in the space right now sort of, what are the major efforts that have already been undertaken and talk a little bit about their pros and cons. I want to be clear that I'm not here to criticize any of these efforts. Many of them were monumental efforts and many of them have done great work that pushes the field forward, but I think none of them you'll see really meet our needs in terms of accomplishing what Seth was talking about and I'll pushback on Seth and say we have a hundred uses for the things that we're creating, we just need to state an implementation plan and work directly with partners who are gonna do the implementations, so that can deploy in the field as rapidly as possible. So, this is an I triple E (IEEE), I
believe, definition of what a data element is. So, there are a lot of different definitions, actually, of data elements and common data elements. I think these are relatively useful, especially the lower one. A clearly defined common data element, "A clearly defined data element that is common to multiple data sets across different use cases." Many of the things I'm gonna present weren't really created for the purpose of multiple use cases and that's a major problem in calling them common, as I already... So, in addition to the requirement that you agree to share the content across multiple use cases and from multiple stakeholders, you also need to have some level of standardization that will allow multiple people to implement the same stuff in the same way. And, as we know, from our experiences in early health IT standards, it's easily possible for two systems that have implemented the exact same quote "standards" and not be interoperable at all. So, for data elements that refer to the format and whatever standard, terminology bindings are really critical. Optionality versus required elements. This is where people get into a lot of stumbling blocks, right? You can say you can add x, y, and z thing, but when you send x, y, and z thing to system S, or system B, system B goes, "I don't know what this is," and it dumps the content and, as I mentioned, use. So, I have the word government here, but the government is not the only person who has created large-scale common data element efforts. Many of these are government efforts I'm gonna talk about, but almost none of them, and I don't actually know if any of them were really intended to apply to direct patient care of the primary use case. And as a person who provides direct patient care still that doesn't make any sense to me because the data that we're generating is coming from a direct primary care use case first, and then flowing down to all of these reuse cases. I like to mention the learning health system over and over again when I'm talking about this subject, but the core concept of the learning health system is that without interfering in direct patient care, we generate tons and tons of data and that data seamlessly filtered out to all the parties who need it, and they're able to use it to do research and to do quality and to do safety and all of the good things that we expect, Health IT to be able to enable and it has not given us that yet. And so, this is, again, partly why because when we have different form and manner and terminology, et cetera for every use case, that means that every communication has to be a one-to-one mapping. Anyone who's done a one-to-one mapping knows that the number one thing you should avoid setting yourself up for is having to do a lot of one-to-one mapping because you never stop having to maintain them, right? Terminologies get updated, practice changes, et cetera, so you basically, for life, are locked into an activity which you hate and which adds only the value of giving you one point of interoperability. So, here I'm gonna talk about, this is for the environmental scan piece, I'm gonna talk about a number of different sets and I talk about these, again, not to criticize them, but to point out that, ideally, we wouldn't have to really create very much content. There's a ton of content out there. All we have to do is throw a lasso around that content and pull it in and create a governance process and get a bunch of people in the room and say, "Let's agree to share this." Now, let's look at it. Is it good enough? Does it bear use cases? If not, how do we harmonize it? So, I'm gonna talk a little bit about harmonization at the very end, but first I'm gonna sort of do a quick whirlwind tour of all of our most, I think, under a couple here that I won't go into detail on, but most of these and what they do, what they're for, where they are. Oh, there it is. So, probably everyone I've heard is either the US Coordinator for Interoperability, which is the new incantation of what was previously the Meaningful Use common clinical dataset, shout out to Rob in ONC. So, this started out as a goal to create interoperability at exchange points of static data for transitions of care and it was updated at the 2015 addition certification rule. I'm gonna show what's in that layer, in part, because CMS has recently announced that they do intend to enforce their requirement that systems are using this version of certification, so those people should have this level of
interoperability starting in January 1st, 2019. The USCDI took the common clinical dataset and turned it into sort of a living, breathing thing where there's annual update process and there is a sort of a taskforce group of experts for it to go to work with the community around how to maintain and extend so content appears. I looked at the ONC interoperability form recently and there was a fairly robust discussion on this topic. Mainly that's how can the community actually interact with this process and is this process designed in a way that can effectively interact with the community? And one of the things that ONC indicated is that they need guidance from the community to know how they can make this more meaningful, but also accessible to the average person because if the only way, I think, to interact with the USCDI, is that you read a several hundred page pdf document and submit public comments that's probably not acceptable to the average clinical person or even all of the clinical organizations. But, it is something that they're working on and I encourage you to give Rob or other folks from ONC your ideas and feedback. So USCDI describes interoperability at the data class level. That kind of makes sense, I think, from their perspective because staying on each individual data element would mean taking on some fraction of 40 million, obviously, a lot of those things are duplicates, but, so they have to take the data class level because that also tends to be the approach that was taken in the common clinical dataset, but in reality, it's actually a combination of data elements. In fact, I'm gonna show you in one second. Because there's some focus on the data class level, there's a lot of content that's not granular enough to ensure interoperability, right? You can say there's a field for blood pressure, but if there's no guidance about range and how the blood pressure is represented, like is the diastolic a separate field from the systolic? Is there a unit that should be attached to it and if so, how? That's a place where two different people can interpret it differently and two different systems can't read it. So, these are the items in the 2015 edition. They are not items from the proposed most recent USCDI, but there's a couple of additional classes there. But, it shows clearly date of birth, a data element, problems a class. So, these things are sort of mixed together and one of the things that was called out as the request for ONC is can ONC create more clarity in what they mean by data classes and can there be a discussion about what level of granularity they want to cover in terms of moving this forward. I'll say, it's not entirely clear. The US, the CCDS was something that supported directly by certification testing. It's not exactly clear how certification testing will support the USCDI, i.e, would there be a new edition of certification every year to support it? I will not suggest that I put Rob on the spot and ask him to answer that now, but it's food for thought and wonderfully, for many registries, the UDI is on that list, so we should see that appearing in every system in 2019, January 1st. I'll also note there's some unstructured fields, which I think, as a clinician, are very helpful to provide some level of system-to-system being able to know where to find things, like goals and assessment and plan of treatment. So, we'll see, coming up, how those actually become usable by clinicians at the point of care. And I'll start moving faster, I promise. This is just a quick slide of a graphic from ONC describing how they expect USCDI to build upon itself every year, right? They don't dump content from year-to-year, they just continue to add to it. And the model, I think, is very useful for us, thinking about how do we create the data elements that we need and then expand that content gradually out until we're able to cover the whole sort of universe of clinical terms and administrative terms that we need. PCORNet. So, PCORNet built on the PCORNet Common Data Model, which was based on the FDA Sentinel Initiative, and people can put that stuff up here. One advantage is that leverages have their own standard terminologies. Many of these efforts I talk about, one of their strengths is that they utilize, or require standard terminology for every term. The output, at least at the PCORNet level, is in PDF and Excel files on their site. It is a research-focused use case, so not focused on the direct primary care use case, per se, directly. But, it is used by this
PopMedNet platform. And, as you can see from the list that I just showed, there's a fair amount of overlap. So, I was whispering to Joe earlier, it would be fascinating to know the 40 million terms that are listed in Jimmy's slide, how many of those represent unique entities and how many of them are the same I see over and over again?

- [Seth] So, one comment here is that these are all classes, right? So, just as a way of contrasting, just with the USCDI and such, where they kind of mix and match the element concepts in the classes.

- [Julia] Absolutely, and there are items, I believe, within all of these and I don't know how death's cause is structured specifically, but you can see, yep, sort of from a clinical perspective, I think it makes sense for us to think about groups of some concepts as classes and groups of other things as specific data elements, but you always have to remember, can a machine understand the difference between members of a class just with guidance at the class level? And, I think the answer's probably no. FHIR. So, FHIR is all the rage and FHIR is excellent for its purpose, Fast Healthcare Interoperability Resources. This is the data element resource. But, FHIR is not a content standard. FHIR never said they were gonna go out and create a bunch of CDEs. They don't have a CDE repository and I don't think that they ever will because like NLM, they don't perceive their job as to sort of act as the police person of content. FHIR interoperability is focused, really, at moving information, not what the information is. And it doesn't have an underlying logical model, so that means as the standard, itself, updates, the content, itself, changes if it's only represented in FHIR. FHIR Core does contain standardized, required supported content, so when you stick to the core level, interoperability is, in fact, pretty easy. Although, again, you run into that question of optional metadata and that becomes a sticking point. FHIR US Core is, in fact, designed to support the CCDS, USCDI that I showed earlier and it does. But, extensions in FHIR, and there's obviously a ton of content in extensions because the core right now is still very small and not cross-tested for conformance. There's no governance process, so there's no guarantee that if you pick up two set extensions in FHIR that they will be interoperable. In fact, there is a fairly decent likelihood that at some point they will be non-conformant with one another. So, this is CIMI. Clinical Information Modeling Initiative. I switched the slide order, sorry. So, CIMI is an international collaboration dedicated to providing logical models to decide content. The content being in a logical model means that that model can be stable even as you can translate the model into multiple different physical representations. For example, FHIR is one of those and it takes the open-inclusionary process of anyone can participate in building content and providing feedback on content, but again, it's not a physical representation, so you need to decide to implement it specifically in the standard at an O-4 kind of system level before you have the ability to use it. So, this is just an example of the CIMI detailed clinical model. One of the good things about CIMI is that it really does take you down to the level of detail necessary, so that there is no optionality. The system will know every piece of information that needs to be there and how it should be formatted, at least at the information level that a clinician would need. The CMS Data Element Library. So, this was presented a couple of times recently, including at the ONC interoperability forums. The CMS Data Element Library basically takes a bunch of long-standing required long-term substitute care assessment data element sets. These are usually a long form that someone might have filled out on the Scantron kind of sheet in the past and it puts them into standard terminologies. There is some structured metadata. It's not using, it's boosting formal standards, but this
is what we talked about, they don't have an API to it. But, you can go and look at the list, and search for one of these elements. One of the advantages of some of the content here is that it's very focused on functional status. So, if you're looking towards patient-centered outcomes, some content here and codes and things might be useful. The CMS eCQM Data Element Catalog. So, this is published annually since 2012 with the CMS Electronic Clinical Quality Measure Set. It is a formal common data element in the sense that it has both a data model, data type with attributes, as well as a value set or a code published with each concept. I will say that across itself, there is generally, there has been some overlap and duplication, even within the set although, CMS has been working to try and eliminate that as much as possible. One of the good things about this, though, is that it's required to be ONC certified in the system, so these concepts are widely implemented in the way that you'll see them in the data element catalog. There's not any other metadata beyond this data set and attributes. There is a formal QA process, though, and so, some of the things I'll mention are things that maybe we have learned from the processes as we're moving forward with a next generation process. And they have these public feedback to attempt to do harmonization and improvements. And again, they're not communities that have CDEs and they're not designed as primary clinical care case, although that's in the mind, I think, of the developers, but they're largely defined as an owner of measures, so again, here's a luxury of creating content that's for your specific use case, your expectation is probably not that it's going to be most super friendly to reuse. This is just a shot because there's a downloadable version of it. Although, you can use the NIH site's API to directly pull this content from a machine-to-machine level. The last system that I'm gonna talk about, the NLM Common Data Element Repository. So, at the recent HSPC PCPI meeting, NLM director Patty Brennan talked about NLM’s effort to try and corral all of the different common data elements, resources that they know of into a single repository. They've been working on this guide since about 2014. And these are the data sets that they've pulled together within it. They've continued to talk about what the right standard sort of manner is and the one thing that they won't, and will never do is provide governance for content because NLM thinks of itself as a librarian. Librarians don't tell you what you can and can't read, they just provide you with the books, right? So, the community will still need to provide governance across procedures they're using in content. So, in conclusion, I'll argue that CDEs are like registries. They can unify multiple use cases. And what we're talking about doing, both at the registry level and at the CDE level, is bringing together the activities around decision support, measurement and analytics, reporting, research, guidelines, all of that into a single representation that supports all the use. This is just a little summary of sort of the approach that a number of efforts have taken to harmonize data elements and I'll argue that this is as good of a starting point as any. I'm gonna start by doing a scan and defining what your scope is, bringing together your stakeholders, thinking about what your standards and workflow requirements are and then you'll want to look around the world, probably want to look at a lot of the places I just showed you, to pull in all the content, all the formats and definitions that you think are overlapping or in scope. And then, you want to do the process where you identify overlap misalignment. You want to say some stuff that doesn't reach a quality bar and so it will be discarded. And the high-quality content is going to be a starting point for us to build out something that meets all of our needs as a use case. And you want that team to be comprised of clinicians and technical experts. You want to make sure that you don't skip over the activity of defining that it works and data flow. So, walking through use cases, not only what are the activities that a person would do in this use case, but what's the information that's generated automatically, what's the information that we ask people to generate as they're working through the process to help you to avoid gaps, and also to avoid asking for information at the point where it
wouldn't match what you created. And then, obviously, you want robust QAs, and lots of public feedback and review and then you can do a release and then you can basically start over with the QA and feedback steps, ad nauseum probably. So, I'm arguing that the CDEs that we're talking about here are designed to close the interoperability loop or gap, right? They will cover, eventually, all the necessary clinical and patient related concepts needed for clinical care, a la USCDI's kind of approach. This, if the clinical and technical community take on this activity as a group, it will remove the burden and the risk of the federal government creating this content for us. The federal government has very good intentions, but they certainly don't have a direct knowledge of the use cases that all of us are experiencing on a day-to-day basis and I'll argue that that makes us a better group of people to provide this governance. It allows organizations to then perform a single mapping, you can dump all of your one-to-one mapping activities. That will save you an infinite amount of time and resources over time and it will allow people, one of the things that patients are always saying is, "Well, I don't understand how you can't get this if my iPhone can do all this insane, cool stuff." Well, it's actually because the content is not standardized. So, a person who's not a clinician right now can't come in and build some amazing cool app in five minutes like they can build a video game because they don't know what content is and how to define it and they have no guarantee that a system will be able to track that information, right? If we can create the building blocks for a developer to come in and start doing this, then people who have never been able to access the health IT sphere, can suddenly come in with their technical knowledge and build the stuff that we really want to see. And so, this is sort of the vision. Seamless semantic interoperability thanks to common data elements. Elise.

- [Elise] I have a question. Can you go back to the CIMI model of blood pressure?

- [Julia] I'm sorry. Stan may get to answer this question.

- [Elise] Okay. Well, actually it's a question for the clinician. So, ultimately, it comes down to the clinician and the patient are sitting in the room and the doctor has to record the information. So, I see that you have a data element for sitting. I mean, are you going to require the doctor to put in that the patient is sitting or do you just assume it is? Then, do you get a missing data problem? How do we, or even the right arm? How do we come up with a practical set of data elements that can really be collected accurately at the point of care?

- [Julia] So, and I'm gonna let Stan answer this, too, but this is my two cents. So, I wouldn't ever shown this to the clinician, generally, right? This is like gobbilily-gook to them. One of the weaknesses of our current system is because we don't transfer information in a directory format seamlessly, we don't benefit from all the documentation of previous users, right? So, over time the gaps should be smaller and smaller for those records, but in fact, I spend a lot of time reentering the same information that's been entered over and over again in the past, right? So, one of the useful things about having a structure that extends beyond what the workflow activity data entry would be is that it allows, over time, you to fill in more gaps. For example, if you put a cuff on a patient in the ICU of this hospital and
they're wearing that cuff and it's there on the arm and they're in the bed, then this information might be automated somehow. And, functionally, as devices are getting more and more sophisticated, there are more and more ways to enter information automatically that currently would be manually entered. That being said, if you look at a lot of the research data elements, is that this is part of the problem, they get so detailed and into the weeds that they're really hard to use and in those cases, oftentimes, that metadata is often considered required because the use case is a person who's not providing direct care is entering the data, right? Stan.

- [Stan] The idea is that in the general model, the idea is that in the general model, almost all of those elements are optional, so you need to have the, absolutely have to have the systolic measurement, but the body location, the patient sitting, they're actually, this model is a lot bigger because you could record the devices being used, you could record the exercise or the degree of exercise that's going on, et cetera. And the idea is that from the general model, you'll have specific use cases and it may be that for in a clinic visit, you just don't require any of that information because they're not interested. But, for instance, if your use case is that you're trying to do orthostatic blood pressures, then you have to know whether they were sitting or lying down when you took that blood pressure. And so, in particular instances, clinicians agree that we have to collect those data points to have the kind of data that we need to be able to provide quality or analytics on that and if you're in a different situation, you're gonna require, for instance, in treadmill or other kinds of cardiology things, you're going to require to know the degree of exercise that's going on at the time of blood pressure and so, you try and tailor it to the specific things, but the point is that some people will collect data without all of the detail, other people will require the detail, but all of that data to the extent that... if you pool it together, you know they're all blood pressures. And so, there's a lot of detail in this and it comes back to another principle, which is there's a part of this that has to do with modeling and how you model data. But, there are other things that we can't accomplish unless we agree with each other what the important data elements are that we need to collect and what degree of detail we need to collect. And so, people come in and we're trying to understand quality about how we treat patients with chest pain and some people collect history and physical and other people do a 12-lead EKG and other people do laboratory tests, et cetera. If we all do different things, our ability to actually correlate and do analytics across that is dramatically diminished. And so, there have to be smart people, domain experts who say, "Look, you can collect what you need for research, but in terms of understanding quality across emergency room visits for people that come in with the chief complaint of chest pain, we need to collect at least these things if we're going to be able to understand quality and understand principles." So, there's a lot more to be said about this. In our situation, one of the big motivating things is that if clinicians understand why you're collecting the data, then they don't mind the burden so much. But, if it just looks like recreational data collection, then people get angry.

- [Julia] Yeah. If you have a hundred million blood pressure readings right now, I pretty much think you can guarantee that you'll be able to extract zero sitting blood pressures because of the lack of standardization. Even if only five million of those blood pressures contain the metadata of sitting, if you were trying to do some sort of research project where you understand sitting blood pressure, then, now you have five million measurements, where you had zero before. And it may be that for quality metrics
on blood pressure control, you'd really have to have the patient sitting and if you're not willing to document the patient sitting, then you're not compliant with checking their blood pressure correctly. So, there are use cases that, you know all of the downstream use cases can be enabled as long as there's a fundamental way of describing the data that every system can handle. Yes?

- [John] John Windle from the University of Nebraska Medical Center and this is very timely 'cause I'm sitting on an ACC-AHA taskforce on performance measures for blood pressure. And this is exactly the question that I'm asking because the guidelines are very clear. You treat people to a blood pressure less than 130 over 80. Thank you. And then I said, "Well, which blood pressure do you use?" And they started saying, "Well, the home blood pressures are different from this", and I actually am very interested in this group helping me to come back with the, I mean, this is, fundamentally, should be the easiest thing to do. It's not chest pain. It's a number. It's a limited field. And there is no guidance for doing this. I was talking with Joe Drozda last night, and for the SPRINT trial, it was, they were requiring three blood pressure checks with an automated cuff done five minutes apart. So, they had their methodology sustained, so that they could say we're doing three blood pressures, then average. That I can understand. Now, when you take this into a clinical domain where my MA checks the patient into the room, as a patient told me yesterday, "Well, my blood pressure's up," "but they didn't wait, they told me they specifically didn't have time to wait for me to sit here for 10 minutes to recheck it." And I'm going, "This is hard." So, as sitting on someone who's supposed to be publishing a document in the next three months, what do I say about blood pressure from an interoperability perspective?

- [Julia] So, I think that that's where the value of the model really comes in. So, if you look at the existing clinical quality metadata on from CMS, they would treat blood pressure and orthostatic blood pressure as just two separate things. Right? It's less for people in the field to figure out what the relationship between a blood pressure and an orthostatic blood pressure is. If the model can provide a definition and relationship between that information, the system can actually better support users in extracting this information without additional effort, right? That's the patient, so the tests, and the test goes off every five minutes, and the test records the blood pressures, and that goes into the system, then the system can know, this is an orthostatic blood pressure, here are the time intervals and here are the three measurements and the patient was sitting with the test in these positions, we can support users with clinical decision support for data entry and not just for did you comply with an evidence based guideline? And we have to do that because that's why people are so angry, right? That's why I get very frustrated when I have to enter the same information that I know is in the system. Why should I tell you that I'm not giving aspirin if the patient has nine platelets? You just told me the patient has nine platelets. This is insanity if you're a clinician. It's obvious to you. The system is telling you, why shouldn't the system be able to understand that? So, there are technical barriers, but, again, having a representation that the machine can understand and that creates a single method of relating information that's related is very important to presenting less interoperability problems we've had with this. And Stan will help you with you later.
- [Audience Member] Can I add something? Thank you. So, I'm a pathologist, I spent a year in internal medicine, so I have that experience and then I've designed systems for collection of tissue for research and one of the things that I learned is you have a data model. This is all the important things. Maybe you don't always collect them, maybe you do. That helps. But, the way in which you collect them has to follow the workflow. Whenever I get my blood pressure taken in the clinic, I am always sitting down because that's where the blood pressure machine is, right next to the chair. The nurse sits me down. The nurse shouldn't have to say, "She is sitting." It should be, you're measuring it here on this machine by this chair, the patient is sitting and it's filled in. And I think, although I know I'm passing the blame onto another entity, is that while we have data models and probably way too many of them, our EHR systems have been, well, we'll just try to collect everything, so that you can put it in your data model the way you want and it becomes so burdensome to the user that it's impossible to use. So, my call out is to the EHRs, is throw away what you have. It's built in 50-year-old technology. Take a look at it. It looks like Visual Basic from the 1980s. Throw it out. Take a look at a smartphone. Every time I try to verify a cyber-genetics report, I have to type in my password again. Really? Why not biometrics? Probably better than a password. There's just so many ways in which the electronic data gathering and recording community is about half a century behind where the big data community is in terms of understanding how to apply medical care and that's not even just at a research level using big data, that's just in being able to utilize what we know. So, end of my speech.

- [Julia] Yeah, and you know, I have mental models of how this stuff works. It's sort of just ingrained in my head. If the community together comes and spits their mental models out into this and then we build machine-readable things from that, then over time we develop a collective mental model of how information exists and is structured and relates to one another. I think it's very valuable in improving the quality of data that we're getting. Right now, when I look at the EHR, I see a field and it's usually defined by a single word and I have no idea what's gonna happen to that information or what it should include and what it shouldn't include, et cetera. If we include in the model definitions of terms, then that information can be populated by the machine, so that if you want to hover over the field, for example, you can see what's this data used for, what's the data mean, what does it include, what does it not include? If you want to see that. Have I been talking too long?

- [Dr. Tcheng] No, you're doing great. So, this has been terrific. So, I'm gonna ask you a question from the registry's perspective. So, part of the genesis of the project here has been what I would call, falls into the category of nature abhors a vacuum. There's all this work that's been done for the standardization of these elements, yet, using the word that you used, which is corral, nobody's corralled all this information together. So, that's in essence, what we're, I think, what we're trying to accomplish here. But, my question for you is now that you've done the environmental scan and you've actually identified the lack of a corraling function along the lines of what we're trying to do here. What do you see coming or what would your recommendation be? Where should the registries be looking to? Where should they be going to try to find this type of information in a way that allows the registries, themselves, to be successful? That's really the focus of our meeting, is to how to enable registries.
So, you know, full disclosure, I mean, the opportunity for most of the people who created the things I'm talking about here was relatively limited to harmonize with other content because traditionally, most of the content has not been easily accessible or electronically created. Traditionally, many of these efforts didn't go through a broad engagement process with many different other use case owners and other kinds of stakeholders. But, now we see that those things really need to be requirements, right? The only organization I know that really is talking seriously about trying to work together to corral content and be transparent is the HSPC CIIC organization. I think that, fundamentally, unless clinicians really get together and put their weight behind a forcing function onto other people creating and distributing this kind of content, that there won't be a change. If just realizing we have an interoperability problem was enough to affect a change, that that already would have been happening. And I think that we can definitely leverage mechanisms like the USCDI, right? We can come together and say, "ONC, we want and need you to support the governance process that we're willing to undertake." "We want and need you to support via a regulatory certification testing mechanism for the content that we're going to create." And then, how can the community, including NLM and organizations that currently invest a ton of resources into mapping content and building disparate data models, how can those resources be redirected to a single effort that will yield content like what we've been talking about? That is exchange interoperability data elements that are driven by the primary clinical use case.

- [Dr Tcheng] One more question.
- [Anqi] I think we have time for one last question.
- [Audience Member] Good morning.
- [Julia] Good morning.

So, my concern is that we have lots of people building clinical data models and CDE sets and so, Stan's mentioned many times over the years that that creates a condition for many silos of non-interoperable interoperabilities and at the end of the road, you diffuse the capacity for interoperability instead of condensing it. If there's no sort of CDI, CDE repository somewhere that's accessible to each organization and each functional unit that's developing these data elements, then you're really lost, I think. And I don't, if ONC can establish a pathway to certification of the domain specific data element sets and clinical models, that would be marvelous because at least there's a pathway to getting those into EHRs and other IT products, but still, there's this missing link at the repository level and I just wonder where people see that going.

Yeah, so as I said, there's so many resources already in this space. If we could spend half as much money and get three or four hundred times, I think, better product. But, I think we want to work with people who have successfully taken steps in the right direction. And when I mention ONC, it's for the lever to get systems to implement them and they do that by requiring testing. So, they set a bar for systems, as well. I think the National Library of Medicine, their CDE repository could easily be migrated or improved to support content that we're building, right? We, as a clinical group, should not take on the activity of end-to-end supporting all of this stuff, right? We need to come together and work with the CTTI Core who rightfully develop and support their piece of this process and work with them to make their process tooling support exactly what we need. And, probably, we'll always need to provide
governance. No one has ever stepped up to the plate to do that, which is part of how we got here. Great discussion. Thank you all.

-[Becky] All right. Great Discussion. Thank-you all. So, lets take about a 10-minute break and we'll regroup and then we'll hear about the data elements from Anne and Mary. So, about 10 minutes.

(break here – room noise)

-[Becky] Okay, well, thank you all for being back and on time. So, let's see if we can't go ahead and get the next session started. I would like to introduce Anne Heath. She is one of the informaticists at the DCRI and she's been helping us with the data element extraction, and then, following Anne will be Mary Williams who is another one of the informaticists from the DCRI. Anne.

-[Anne] Thanks Becky. And now if you'll notice, if you look at the agenda, Mary and I switched places, that's for logical reasons. So, I'm trying to present some of the background, which will set the stage for what Mary will discuss and then later on this morning what Doctor Tcheng will go over. Before I get into, sorry. Is that better? All right. I'm short. There. Before I get into any of the specifics, I first want to explain, as you recall, Aim 1 of the project was to extract data elements from registry sources that were submitted into us. And our original assumption, the original charge that was given to me many months ago was to find, to select the data elements across all of the sources that we received that were common among 50% or greater. And we thought we would find a bunch of those and we thought that those could then be moved forward as a proposed standard. What we found was something very different. We found that, essentially, zero of the data elements were exactly common among 50% or more of the sources received. And, in fact, very few of the elements were collected in exactly the same way in even more than one source. I think there are a lot of reasons for this and we'll go through some examples. I think one of the biggest conclusions I came to was that each of these registries were developed for very different use cases, for very domain-specific use cases. And roughly 50% of what we received were data models, actually, that collected only their domain-specific data and I'm assuming in there use would be later integrated with the rest of the nation's data. We received about 50% real case report forms. Those were the easier ones to extract because those were specific questions and specific permissible values. But, even among those, they were largely domain specific, so you'll see here in a minute that even what you would expect to be very common data collected blood pressure rates. Height, weight were actually very sparse across all of the thresholds that I looked at. So, if you joined our WebEx about a month ago, we went through some samples, sex, tobacco use and lab. We've expanded that a little bit, I actually, behind the scenes here, have pretty much all the data on everything. But, I just wanted to give you a sampling today of what we saw across these concepts of interest. And, like I said, my hope is that what I show you, in terms of the lack of concordance across the sources, will set the stage for what Mary will present in terms of the national model and then, further, what Doctor Tcheng will present in terms of recommendations and path forward. So, not to belabor that too much, I just wanted to give you kind of the background on how we got to where we are now. Okay,
so jumping right in. What I will be showing you are what we're calling concordance tables for each of these concepts of interest, and race is actually on two slides. Among all of the concepts that we looked at, race is by far the one that is collected the most often across all the sources. But, even then, there's a great deal of variability in the way the data's collected and we counted something as non-concordant if the data element name was different and/or the list of numbers for the values were different. So, if you think about, or when I think about back in the clinical trial world, the main five categories that you most often see, or, here I go, that you would think you would most often hear and see that. White, black or African American, American Indian or Native Alaskan, Asian and Native Hawaiian or other Pacific Islander. What we're seeing here is that a lot of, that a lot of the registries took those five and so, okay. Oh, okay. A lot of those registries took those five and added a few others, like ‘refuse to answer’ or ‘other’ or ‘unknown’, thereby making it inherently, then, not directly interoperable with the primary file. So, I know that this is a lot of little words and you're not expected to necessarily review this in detail right now. So, jumping forward now to ethnicity. Again, this ethnicity was probably the second most common element that we saw among the sources and just like I said with race, a lot of the registries are taking the main two categories and adding additional options, thereby making them not concordant. Date of birth. I expect it's probably surprising to see that, what do we have here? We have nine, 10, 11, 12, 14, so date of birth was collected or it was, yeah, it was collected in only 14 of the 38 registries. Now, that sounds really surprising, right? I promise you I have gone back and re-looked at all of this, and again, to my earlier point that so many of these registries are domain specific and, again, I have to assume that they're intended to be later integrated with the rest of the patient's data, so even date of birth isn't always collected. And I'm just gonna go through the troubles really quickly now because I'm just gonna keep repeating myself. Blood pressure, we've got four. I combined pulse and heart rate just to get a little more bang for my buck. Height, still not that many. Weight, a little more. And we were specifically looking for UDI and I found UDI in only one of the registries. I feel kind of, I really honestly feel kind of like a buzzkill standing up here and showing you the fruits of months of labor, but I do think it's been a really valuable effort because I think the time did need to be put in to demonstrate that even my continued assumptions that there would be a lot more concordance just weren't true. And, again, this goes directly to Doctor Tcheng's point earlier today and I'm sure he'll reiterate this. Again, in terms of collectively needing to put forward a proposal about how to try to keep this from happening. I think, really, that's all I have and I didn't fill up all my time. Following this, so, I have a great deal of data not in the slides. I have a great deal of data in spreadsheets, and I'm gonna load it directly in an Access database and create tables and graphs that will be more shareable and more consumable. So, those things will come in the coming weeks. Any questions?

- [Male Audience Member] I have a question. How many of the registries were EHR based?

- [Anne] I don't know the answer to that.

- [Audience Member] I think that would be important.

- [Anne] Okay, I will take that back now. Yeah, I don't. They were, again, I think they were intended to be either integrated with the EHR or downstream or are you suggesting pulled directly from? Okay, let me take that back.
[Male Audience Member] That's the vast majority.

[Anne] Pardon me?

[Male Audience Member] That's the vast majority of clinical registries.

[Anne] Okay. Somebody have a question up here?

[Female Audience Member] Can you explain, this is Marjorie Rallins from the PCPI Foundation. Can you explain domain specificity? When you say...

[Anne] Right, I know I said that a lot. What I meant was specific to the cancers. Specific to cancer, specific to, I'm sorry, I'm just drawing a blank, specific to any one of the areas of care or specific disease being studied.

[Audience Member] So, the data elements that you just showed, some demographic-type data, race, ethnicity. I'm not necessarily seeing domain specificity variation there, but I could see it in other data elements as you get more clinical and I think that's the challenge, is while you might, a data element might be named like diagnosis or whatever, within a particular context, the data element could be understood to be different. Even if it's common across other domains and I think that's the crux of the issue.

[Anne] You make several really good points that I want to say something really quickly. What I may not have explained well was that what I'm showing are those that are not domain specific. These are basic patient data. And that is what we are looking for. The fact that so many of the sources were so domain specific and we weren't looking for those. And so, therefore, there was very little commonality across them. So, I think I failed to explain that well.

[Dr. Tcheng] So, let me just amplify a little bit about, Anne, what you just said here. The premise of the grant was to identify concepts, I'm going to avoid the word data elements for the moment, identify concepts that were shared across registries. So, patient identifiers, date of birth, sex, race, ethnicity, tobacco, smoking status, we anticipated things like labs, medications would appear, not necessarily the same labs or the same medications, but as a class of things that registries, in general, would be asking for these types of information. The reason for the emphasis, for example, and Julia's talked about the USCDIs, we anticipated there would be a fair amount of overlap with the USCDI. The USCDI, CCDS, USCDI was the predicate work, at least some of the predicate work for this, as well. So, when we put the grant together, we were envisioning it as the first step of a long journey, all right? So, you know the old saying from the Chinese philosopher, "The journey of a thousand miles begins with the first step." The first step here was to identify a place where the registry community could come together and say, "Yeah, we have these core needs, this core need for commonly representing," excuse me, "representing core concepts in a common way" "that if these are physically built this way in the registries then they would automatically become interoperable." So, that's the genesis of the grant. We thought we were creating a very liberal set of criteria for identifying these common concepts. And that is that they would appear
in half of the registries that we solicited. It is, therefore, quite a bit of a surprise that actually very few meet that 50% criteria. Maybe I'm looking at this from an American College of Cardiology perspective 'cause that's the registry environment that I grew up in, but from a cardiology standpoint, how old you are, what your sex is, what your cardiac risk factors are, what your medications are, all those things are important, so that's the framework and I admit that probably I was a bit naive, but that was the framework that I was coming at. One way to think about this is that we're actually done with the grant because nothing meets the 50% criteria. However, I'm not gonna, I'm not gonna let everybody off the hook that easy, including our team here. So, and you'll see what we're going to propose as a result of all this work. One of the proposals actually is to exercise the PI prerogative and say these are the things that we're gonna develop as concepts for dissemination, so that as registries are either brought on or as they go through a revision process that you can adjust your physical, the way you represent the data physically in the individual database. All the registries then will get that much closer towards the interoperability goal. Now, this is only, at the end of the day, this is only going to be 30 or 40 or somewhere around there, maybe two dozen data elements. And, if you remember, I proposed the number 40 million. So, this isn't even .001% (point zero zero one percent). Or, maybe it is (.001%) point zero zero one percent, but this is just a tiny little bit, but it's a step in the right direction. Stan's here and I'm also part of HSPC CIIC. Frank is, Frank Opelka's in the back there. Part of this, this is the kind of a framework that we're thinking about to help move all this stuff forward, but you gotta start somewhere. So, this is the beginning of that journey, so full color commentary. If I could ask, and I'm not sure if we asked the question about EHR-based registries, I wasn't sure I understood the, yes. If you wouldn't mind just explaining a little bit again, what you're intending to accomplish with that question?

- [Audience Member] Jimmy, what I was intimating is, what were you looking at? Were you looking at the data dictionary? If I reversed specific registries, the IRIS registry maybe, but we collect all that data. But it's in the data dictionary, so I'm curious as to where were you looking for the data elements to look at commonality?

- [Anne] The easiest answer to that question is that I was looking at whatever it was that I received. In some cases, what I received were the case report forms and those were the easier ones to cull data elements from because they were typically more complete, I would say, in terms of the rest of the patient data. About 50% of what I received were data models or data dictionaries, which I would also cull for everything. And that's how I had to make belief in the assumption that where there was very little of the demographic or basic vitals or that kind of thing, that that set of data was to be, either was being pulled from the EHR or was to be later integrated with the rest of the patient data.

- [Dr. Tcheng] Exactly. We were integrated with over 50 different EHRs with many, many different ways that the data's represented.

- Right.
- Sometimes we put it into the registry through a mapping process, so those data elements are specific in the registry. That's what I meant by it's not EHR based, it's EHR extraction.

- Extraction.

- A lot of variability out there, we put it in a common, in a single data format.

- [Anne] And that's why it was helpful when I received, and I actually recall looking at your registry and what I received and I could clearly tell what you were, and I wasn't sure if you were extracting or just collecting a small bit and then later integrating, like I said, but it was helpful to look at that data model or just that data dictionary, but what I didn't see, and maybe it wasn't yours specifically, but what I didn't see often were all of those other data elements that would otherwise exist in the EHR, but weren't specific to your registry or that domain of study. Did I say that all right?

- [Dr. Tcheng] Yeah. So, there is a bit of delusion of grandeur here, too. So, you've actually identified the problem and that is if you have 50 translations to get that data element into your registry. What if we turn that equation around and said all of the spreadsheets, now, I know we only had 38, but there's somewhere between 80 and 100 major registries out there, but we got all the registries to say we always want data bursts in this format, okay? Now, you're pushing pressure on the EHRs and the HIT vendors to store it in exactly the same way, so that you eliminate that translation layer. So can we come up with a core set of data elements that spread across the universe of the registry community that then puts pressure upstream to the EHR vendors to do exactly the same thing.

- [Audience Member] Where does this pressure come from? The pressure comes from we do, just like Medicare Advantage this week gave stars, we've taken our 10 highest EHRs for use, number of physicians, patients covered, and actually we develop a common, it's kind of a bottom up approach. That puts pressure on the other 38 to come along 'cause then there's a huge turnover in clinic, clinicians' EHR work now, so we actually will publish which ones that get a gold star.

- [Julia] So, I snickered a little bit when you showed the variability in race and ethnicity because a lot of... OMB very specifically says exactly what should be collected for those two categories. So you would think if there's harmony anywhere, it would be there and the fact that there's that much variability there, I think is just a testament to how difficult it is to corral this kind of stuff.
- [Anne] Yeah, you just said it better than I, but that was exactly what I was thinking because prior to my role in informatics, I spent 10 or 12 years in cardiology clinical trials and we're very accustomed to the specific five race categories and the specific two ethnicity categories and there was no opportunity to alter that. So, I, too, was, I guess I snickered, I was like, "Wow, if we can't even have "commonality on those, then, I agree."

- [Matt] So, I'm Matt Elrod with the American Physical Therapy Association.

- [Anne] I was like, where's that voice coming from?

- [Matt] So, a couple of things to piggyback on some of what Julia was talking about. How many registries actually did piggyback? 'Cause I saw you had for race and ethnicity or whatever for the common data set, do you have that number of how many actually did do that?

- [Anne] I'm not sure what your question is, I'll have to go back to that slide.

- [Matt] So, basic, yeah, so Julia gave her presentation, she was talking about the ONC's common clinical data center and how there's very specifics about what they say for race and ethnicity. Did you go back and compare how many actually did that?

- [Anne] Well, Julia, if I'm correct, it's just those five at the top.

- [Julia] Right.

- [Anne] Okay.

- [Julia] Number four is the one that's the line of is this being required from all patients.

- [Anne] The first one up there, I counted as not concordant because the question was asked differently. But, yes, it's the five race categories that you see with the concordance count of four. So, the answer to your question is four. Out of 38.
- [Matt] Were consistent with the common clinical data set?

- [Anne] Yes.

- [Matt] That, then that makes me, oh, okay, so that's one, two, three, four, the bottom down. Okay, great. Next question is, how many registries do we know are using data standards for sharing their information?

- [Anne] I don't know that answer.

- [Dr. Tcheng] Yeah, so the answer based upon this work is none. That's the opportunity here is that everybody kind of adheres to data standards, but not explicitly. Definitely not at the level of moving data around.

- [Female Audience Member] When you say standards, do you mean like an API or specific?

- [Male Audience Member] Like the CCDA standards.

- [Dr. Tcheng] I'd like to comment on that briefly.

- [Seth] Seth Blumenthal again from AMA. So, just at the level of content, I think in a data dictionary, ideally, if the registry data model uses semantic standards, which I think is what we're talking about now as opposed to transport or other layers of abstraction. Then, ideally, that data dictionary would reference original sources. So, for example here is how we define race in this registry and its coming from x, y, z standard. So, I think what Jimmy's saying and what the new team has found is that really there's no, there's none or very linkage back to standards, so that doesn't mean that some of these registries made decisions totally irrespective of the standards, but maybe it's not a well-documented link and so I think that's what we found is that registries are sort of making the best decisions they can and, as you can see here, there is a lot of similarity between these sets, they're not exactly the same, the values sets, but certainly you can see the commonalities, but yeah.

- [Male Audience Member] So, I would say that that is somewhat, there are registries that use semantic data standards, but what we don't have is a preponderance of registries that share a common data standard is the bigger problem, So, we aren't here as representatives of data standards organizations,
we're here as representations of the societies, of the specialties and so, there's no singular voice coming from the registry community to either vendors or governance or anywhere else right now. I think that's the bigger problem.

- [Male Audience Member] You're talking about 40 million data elements out there washing around in the sea and then this smaller set of common data elements that are fairly common across all clinical endeavors. And then there are areas of overlap across specialties or procedural groups or so forth where the same terms can have different meanings and where there isn't an effort to align or even a knowledge that similar elements are being used in different ways by different groups. And you have one body of data here that can help to illuminate those areas of overlap, as well, that are not part of that core set, but are more part of the domain-specific set, but the areas of overlap in the domain specific sets are opportunities for disharmony and lack of interoperability, more sharply divided, I think and so, it would interesting, if possible, to extract those areas of overlap if for no other reason than to communicate it back to the groups who submitted that, "Hey, your group submitted this and so did these others. You might want to chat with each other."

- [Becky] Great, I think if there are no questions, we're going to turn it over to Mary Williams from Duke.

- [Mary] Good morning everyone. My name is Mary Williams and I'll be talking about Aim 2 of this project and, kind of facetiously, I'm bringing you more good news. So, Aim 2 to characterize data elements from the national models. I refrain from calling them standards because you'll see in a few minutes, they're, I'll tell you because I am in DC, they're going along party lines. CCDS, CCRF, PCORNet, Sentinel are all kind of doing similar things, but not entirely the same. And, again, here the, I guess, the big models that I was looking at, CCDF, CCRF, FHIR, OHDSI data set standard, PCORNet Common Data Model 4.1 and Sentinel's version 6.02. And I have FHIR in here as a national standard and I will also follow up each of my slides about concepts of interest with the FHIR recommendations 'cause I thought that was important information to have if we move forward with this exercise. For those of you who were on the webinar last month, we talked about some of the sex and lab results, so the concepts of interest in red are what I will be showing today and they're exactly the same ones as Anne showed in her presentation. So, probably be saying a lot of the same things because I found a lot of the same things that Anne did find. And because I looked at, I had to do this graphically. Pew concept of interest in the center and how did each of these national models represent that information? And you'll see that there are similarities, but there are more differences than similarities here. So, I start with date of birth and CCDS, CCRF define it similarly. It's a derived variable in OHDSI, so it takes the individual components, month, day, year, and combines that into a derived data element. FHIR it's its own thing. PCORNet and Sentinel actually did this the same, so, again, party lines, they do things similarly. Now, looking at the FHIR recommendation for date of birth, it's actually not a must have, at the bottom of this screenshot. So, moving on for race across the national models, and I found exactly the same thing, that neither one, well, CCDS and CCRF are collecting it similarly, but the other national models are doing different things in terms of adding additional options, different flavors of null. I noticed the OHDSI Permissible Value list is blank. That permissible value set was not available...
- [Seth] It's right here. Here in the, this is just an example, so the first two rows from the top, right? So, you've got CCDS, CCRF, and FHIR, so it looks like the permissible values are the values that are the same? So, is it the difference in the optionality? Because you've got the same set.

- [Mary] It's the difference in the data element name, so what they called it. Race, structure, definition. You record race and then race here for the third row. It's just the different options. So, I was looking for exact concordances. If they have the same data element name and the same permissible value I said they were a direct match. If not, then I said they were different.

- [Seth] So, just to clarify what you're presenting here. So, in the case of those two, you have the same permissible values, but different optionality? Or the same optionality? Like in terms of binding or maybe that doesn't really matter in this case, 'cause you're going to select, but, obviously, the data element's names different.

- [Mary] Correct.

- [Seth] Okay, thank you.

- [Mary] No problem. Any other questions? So, again recommendations for race based on FHIR, US Core. Again, so race is a variable that they would include in addition to you must have the patient name, gender and some sort of patient identifier.

- [Anne] From the previous slide.

- [Mary] Uh oh, let me go back to it.

- [Anne] Okay.

- [Mary] Sorry. It looks like in the third and fourth option down, the--

- [Anne] PCORNet and Sentinel?
- [Mary] Yeah, those standards they've recoded, those terms using their own local codes, right? Are the first two utilizing codes that are actually members of the value set? I mean, I know that CCDS should and probably the FHIR is, right? So, those are codes from, right? Let's see.

- [Anne] They are and I know it's really small. The OMB race category. You see the highlight here?

- [Mary] Yes.

- [Anne] So, these are the categories.

- [Mary] Okay. And they're from OMD. I see you're reading, Julia. I was giving you time to finish. So, moving on to ethnicity. Again, we're seeing the same thing. Different models are collecting this information differently. Again, CCDS, CCRF are collecting it exactly the same, including the same data element name. And what was surprising to me here was because PCORNet and Sentinel actually collected it differently and PCORNet was built off of Sentinel, so that was a bit surprising to me, but again, you'll see that there really is no concordance across the national models that are national models. And, again, the FHIR recommendation for ethnicity is the OMB ethnicity category for Hispanic or Latino, or not Hispanic or Latino. Moving on to height. Not all of the national models collected height, which wasn't necessarily a surprise. But, again, across party lines, CCDS, CCRF, and FHIR collected it the same way and then PCORNet and Sentinel collected this information the same way. FHIR recommendations. You have to have a category of vital signs and then the LOINC codes for body height, and the next several of these vitals will look very similar because this is how they are collected in FHIR. The next example is weight. Same information, same party lines. And then, again, with the FHIR recommendation, you need a category of vital signs and the LOINC codes for body weight. Heart rate. Not everybody was collecting heart rate, surprise. And, again, you need a LOINC code for vital signs and a LOINC code for heart rate. Systolic blood pressure, again, party lines. CCDS, CCRF, FHIR collected it the same way, PCORNet and Sentinel collected this information the same way. And moving on to the FHIR recommendation, again, one code for vital signs and then that you're collecting systolic blood pressure with the proper unit. Diastolic blood pressure, same thing. And then, again, with the FHIR recommendations, vital signs, LOINC code and the diastolic blood pressure. LOINC code with the appropriate units of measure. And then, finally, the unique device identifier for implantable medical devices. I've added this information. Four of the models collect this information. None of them collect it the same. Oh, I'm sorry. CCDS and CCRF collect this information the same way. And then, again, the Unique Device Identifier under FHIR, just asks you must have the UDI string to classify this data element. So, that's the end of the examples of data elements that I have for you. I wanted to just give you a visual on how this all works. The registry CRFs that Anne reviewed and annotated will be combined with the national model information that I have here and we'll use this information to inform the list of
recommendations that Dr. Tcheng will be talking about a little later. And that's all I have. Thank you. Any questions?

- [Male Audience Member] You know, it's sort of interesting, organizationally, in terms of what, how Sentinel and the PCORNet data models developed. As you said, PCORNet was based on the Sentinel common data model, but PCORNet did that kind of on their own. And now they're actually, they're actually being coordinated by the same center and the same team is managing both. So, I think the fact that they haven't been unified in terms of the differences you described speaks to the difficulties that people have when you have the same team they can't even reconcile.

- [Mary] I agree. And I'll admit that a lot of the differences that I thought were the additional flavors of null or other or not answered, that type of thing. Maybe the PCORNet data model is trying to, capture every case scenario in terms of any possible answer because you have the races and then you have other, asked but not answered, unknown, or refused to respond, something like that, as a category.

- [Male Audience Member] Here's actually another question about the difference between the two. Sentinel grew up primarily on the insurance claims data side. The Sentinel distributed data network is primarily insurance companies. They've sort of added HCA later on. And PCORNet is actually hospitals and clinicians then coming from electronic health data. I'm wondering if that doesn't have something, some impact on the definitions.

- [Mary] Possibly. I would imagine that you could get any flavor of an answer for any of those questions and I think PCORNet common data model was trying to have a bucket for each response and I think in doing so and with the working groups that were developing with the data elements, probably incorporated them for that very reason.

- [Audience Member] Yeah, my team on insurance claims data, you're not gonna even have the opportunity to know a patient refused.

- Right.

- All you know is what's there.

- What's there, absolutely. Do you see that? Ok. And next, Dr. Tcheng will be talking about the list of recommendations.
Great. Thank you very much, Mary. So, what I hope we've been able to expose to you is a couple of things that I would actually draw as conclusions already. I was a little bit harsh in my previous answer when I said there wasn't any compliance, if you will, with national data standards by the registry. That's totally not true. It's not done in isolation. The value sets, especially, are not created in isolation. Really, my comment was that, at least in our evaluation of the electronic material, the case report forms, the data dictionaries, the data models available from you, from the registry community. When we asked the question, "All right, is this done by a binding to a structured ontology, an ontologic representation of, for example, race, is that how the information is captured or is it done by hand typing, if you will, or some other mechanism?" The answer virtually always was more of the latter than the former. It's not moved, if you will, electronically from one system to the next, especially via, if you will, data standards. At Duke, at least, we have a number of our registries, a number of registries that we didn't participate in and a number of registries that we contribute information to that we've actually hand coded, that's the way I could describe it, hand coded a process for capturing the information at the point of care in ways that then get packaged and then submitted to various registries. But, that still is not done as an exercise in true interoperability, that's done via a one-to-one linkage. Getting us closer to that goal where I think we all need to be, but still not addressing that. As best I can tell, we were probably as far along in that journey. We have the advantage of having a number of the registries housed across the street, so we could just kind of walk across the street, and say, "Well, how would you like this data?" and, therefore, get some of the answers. But, suffice it to say that we still are a long way to go, I think everybody recognizes that and actually that's what we're trying to address here, is how we move away from the old models to the future. So, with that, let me just walk you through a little bit of intro slides first, just to set the stage, and then I'm gonna ask you to put on your thinking caps because there's a couple of queries here that I really need your input on to make sure that we're going to deliver, if you will, the best possible product. So, we're here largely as a representation of the registry community combined with a number of our fellow colleagues 'cause they're very interested in leveraging, especially, for registries for various purposes. We have robust evidence. This is from the American College of Cardiology, but it's true for really any of the professional societies, professional associations where part of the business of the associations is to take evidence that's been generated through clinical files, assess those lines of evidence, bring them together, aggregate them, create guidelines, and then, to hopefully implement them into clinical practice. So, if you ask the question, why do we have registries? Kind of the starting point. The real answer is, perhaps, best illustrated in this phrase Anne borrowed from my ACC colleague, "Science tells us what we can do; "Guidelines what we should do; "Registries what we are actually doing." This is a clinical focus. This is a clinical use case. We want to know what's going on at the level of care, the level of provision of patient care that then illustrates or tells us how we're performing, what our outcomes are, what our performance measures are, what our quality measures are, et cetera. And so, if you wanted to think about a use case where it is absolutely critical to capture that information at the point of care in the way that then gets seamlessly and semantically accurately moved over to an assessment tool, that is, registries, this couldn't be a better way of doing it. I showed this slide before, this is kind of one of my fun, favorite slides. In our surveys, and this was part of the background work that, especially Anne, I asked her to do was to see how much of this was actually going on versus true point of care capture of data and our best estimate is that it's in the high 90s, that this is how data still today gets submitted into registries. There are a few exceptions. I mentioned the Duke exception where we've hand coded some of this. You could argue that
folks that are data aggregators who have tools where they actually then query the electronic health record are accomplishing at least a better version of electronic data interoperability, than swivel chair interoperability. But, even there, it was translational error. What's happening is that you get the tool that goes in and sorts through your electronic health data and pulls out specific pieces of information and then moves it over, as opposed to actually, again, capturing it at point of care for direct submission. This is what we've tried to accomplish at Duke. On the very left-hand side of this is the health information technology and other EHR systems that are used at the point of care to capture information. It is semi-structured. It doesn't do away completely with text. The richness and robustness of text is still respected. But, a fair amount of our information, especially for the ones who are going from left all the way over to the right works for the registries is accomplished through structured and semi-structured query. That is, at the point of care a form is presented and the nurse, the physician, the clinician, the tech, whoever it might be, is asked to follow the form and then do the work via the form. That doesn't mean it has to be filled out in the form. So, for example, in the cardiac cath lab, where we're doing cardiac catheterization procedures, what is typically documented by our staff is the play-by-play. All right? This happened at 10:24, this happened at 10:25, this medication was given, this guide wire was used to cross, and then we take that information either as data or sometimes we parse the text phrases that are captured and then bring that data across as discrete data, that is, as CDEs. That way we create structured documentation and then we can use it for the other purposes. Why did we end up, or how did we end up in our current state of interoperability? There's a lot of protean reasons. I promised you I was going to insult a whole bunch of folks, so now I'm going to insult the HITECH Act and ARRA, nothing personal. Marjorie, I know you were part of the work that was put together, but I did want to bring forward this particular slide because I think it's very telling as to why we are where we are. And that is if you look at the work of Jamie Ferguson and John Halamka, and the HIT Technical Committee, in terms of the standards for interoperability, they took what I would call the politically astute approach and argued that interoperability needed to be accomplished between entities, but they really did not want to be in the business of specifying at the native database level, the construct of concepts, et cetera. But, in terms of moving them from one place to another, that is at the boundary between entities, that's where they wanted to work. Now, the problem with this is that we let a thousand flowers bloom. Everybody is doing it their own way, including, if you will, the members of the registry community. That is, if you want to capture date of birth, some put DOB or date of underscore of birth, or date capital of capital birth, whatever it might be. You're free to do that as long as then you can translate that when it comes time to moving that data from one place to the next. It's that data translation, though, that now has burdened us because as we want to move from point to point, to point to point, to point to point, and you get the idea that now we have the difficulty of having to do translation every time you want to move data from the source over to any particular consumer. And so, the question is, again, how do we address this? And so, this is a message that we've been trying to work on here and that is think about it at the level of the native data interoperability, and for all of you registry owners, this is probably fighting words because you're saying, "Wait a minute, is he up there saying that we ought to do it a certain way?" In fact, I'm suggesting to you is, actually the answer to that is yes. We've tried it the other way. Now, let's see if we can move forward in terms of accomplishing native data interoperability, such that we have a lot less work in terms of translating terms. All right, so how do we accomplish this? It is through data standards. Everybody's seen this. I hope you get a little bit of a chuckle out of that again when you read that. But, this is what we're proposing. So, we are gonna, or we have, I should say, started off with this concept that we would take a
whole bunch of registries, by the way, the original grant suggested 20 and by the time we got to 20, we said, "you know what?" "We gotta actually expand a little bit more." So, we took it on ourselves to invite as many registry owners to the party as possible and we ended up with 38. And then Mary and Anne, especially Anne, had gone through all of the case report forms and the data models, et cetera, that look for common terms and we've also tried to match that or marry up to what's already out there, starting off with the USCDI. So, to just make this just a little bit more readable, the USCDI has 22 concepts. Some of these are very discrete. Date of birth, sex. Very, very discrete terms. Some of these are classes of things. And some of these are even worse than classes, I would call them very amorphous. And so, the things that are classes are, for example, are medications, lab results, et cetera. Those, you can create a framework around those, around which that type of information ought to be captured. Types of information that are probably very amorphous: assessment and plan of therapy. Why do I say that? As a physician, I can't think of two patients who might have exactly the same plan. And so, if you think about how that then translates into, excuse me, discrete data elements, that's almost impossible. How are you gonna take 10 thousand patients by taking care of them in the past decade and each one with a different plan and now create a categorical approach to capturing that information? So, I don't have any aspersions to being able to treat or fix all this stuff, but then ask the question, what is relevant at the level of registries? Again, remember this grant is around registries. And we've come up with this list in terms of what's relevant to registries. Alright? So, patient name, date of birth, sex, race, ethnicity, smoking status, lab values and results, not necessarily the lab tests themselves, we're not ordering testing registries, but the results are important, vital signs, medications, care team members, procedures and UDI, probably the most important thing is not what is left, but what has been removed, at least from the USCDI. So, we didn't find any registries that asked about what preferred language, so we dropped that. Again, I made a comment about this, that registries aren't ordering lab tests, so we took that off. Problems. You might think, "Why did the problems thing "come off the list?" Well, problems is going to get resurfaced, not as problems, but as risk factors. So, from a standpoint of thinking about registries, typically, the registries are already predefined. I had a heart catheterization and an angioplasty, therefore, I'm in this registry. I had a hip replacement, therefore I'm in this registry. I had x, y, z, I have x, y, z disease, cystic fibrosis, therefore, I'm in this registry. So, they're already well defined in terms of the entry criteria into the registry. Medication allergies, same concept. It's not something that appears in registries. Registries, they're not really interested in allergies. Obviously, something very important, clinically. Very important, clinically, but much less relevant from a registry standpoint. Immunization, obviously there are immunization registries, but that wasn't the focus of this. We actually didn't solicit any immunization registries, so maybe that was selection bias. Provenance, goals, clinical notes, assessment, plan of care, health concerns, very amorphous concepts. So, we did want something that we could actually deliver on. All right? So, query number one is what do you think about this list? So, as is from the USCDI on the left-hand side, patient name, date of birth, sex, race, ethnicity, smoking status, procedures - procedures are relatively straightforward because CPT coding and the UDI. And then, on the right-hand side, there are seven other CDEs that I would call modifications of what's in the USCDI. So, for vital signs, it's height, weight, blood pressure and pulse. Lab results is actually a modeled concept. We talked about that at the Webinar. I'm not gonna belabor the point today other than, again we have this model for bringing in lab results. I am going to ask Tom Windle to talk a little bit later about the medications model that he's working on. I think you'll be pretty amazed with the progress that he's made. Out of the members of the care team, the only one that seems to be important, at least from the standpoint of registries, is the physician operator, especially physician
operators who are doing procedures, like putting in devices. The rest of the care team, the nurse, not that everybody else isn't important, but from, again, a registry perspective does not seem to be all that relevant. And then taking the investigators prerogative, as I said, these are other things that I anticipate seeing in registries, alcohol abuse, substance abuse, and it would be nice to know what your vital status is, At least getting partway down that journey. So, this is the list of common CDEs, core CDEs that we propose will be the final work product of this particular grant. I'm gonna stop here for a moment and just ask for comments. Frank.

- [Kevin] Kevin.

- [Dr. Tcheng] Sorry, Kevin. I'm Sorry, Kevin Baskin.

- [Kevin] That's all right. Where is infection, death, pain?

- [Dr. Tcheng] So, death's there.

- [Kevin] The outcomes that are important to the patient and goal forming for clinicians.

- [Dr. Tcheng] Yeah, so the answer is we do not have that as part of this particular grant. We haven't done it, kind of on purpose. I wanted to take some fairly straightforward discrete things that could be then modeled, et cetera, et cetera. The rest of that story, and I mentioned that in the very, one of the very first slides that I showed is that AHRQ is actually working a lot on outcomes. They're talking about both patient reported outcomes, as well as clinically-relevant outcomes, pain measures, infectious disease, et cetera. I didn't want to, purposely, I didn't want to step on the work that they're doing, so Elise, do you want to comment about AHRQ and what you guys are doing with outcomes measures?

- [Elise] Yeah, so we, for a lot of you who don't know me, there's a harmonization project across five different clinical areas and we actually have our final report is up for public comment right now. And, but then, there are also other groups at AHRQ working on standards for PROs and so there are a lot of work in that area. But, there's also a lot of work in all of these areas. So, with the lab results, are you working with the FDA project? The, I forget what it's called.

- The SHIELD Project
- The SHIELD Project, yeah.

- [Dr. Tcheng] with Mike Waters, yeah. In fact, Mike is not here. I didn't see him, but he's the author of the lab inspection for this, yeah, yeah. So, there's collusion. There is convergence and collusion as much as we can, but Kevin, to answer your question, I'm gonna leave you in Elise's good hands to try to figure that one out, okay? So, the grant itself is limited, all right? It's a good start, but remember this is what I would call the first steps on a long journey where there are multiple people on this journey, this just seemed to be the area where the greatest opportunity to bring the registry community together existed. So, that's why this looks. Elise, yeah?

- [Elise] I just have a, I just have a question. So, one thing that came up in our project was the issue of medical marijuana and, as the government, I think we weren't really allowed to say that, but as the private sector, I wonder, it seems like it's very important because a lot of people are using it and there's no records and so, when you have alcohol use, like we were talking about related to pain, people were self medicating with alcohol or medical marijuana and also, illicit drug use related to the opioid crisis. So, I'm wondering since you did add alcohol, whether you can add those things?

- [Dr. Tcheng] Yeah, so that falls under the category of substance use and so, since I should be politically appropriate, I just took the magic of power point it is took Abuse off, so yeah, that is something that needs to modeled. Interestingly, in the study of coronary diseases, this is an aside, but in the study of coronary disease, marijuana smoking is a very important risk factor. In fact, it looks like it's, if anything, it probably depends upon which study you read, but twice as effective at causing coronary disease as tobacco smoke. So, in other words, it's twice as worse for you, but, or two times as bad for you, but I think that is important. That's one of the things that we're adding to the listing. Anita.

- [Anita] You just fixed it. I was going to say it should be probably 'substance use' rather than--

- [Dr. Tcheng] Okay, good. Alright.

- [Lou] I'm Lou Meyer with the American College of Radiology and I'm looking at those eight elements and seven of them seem quite clear cut, but the one for procedures, it seems like there's an unlimited number of procedures across specialties and so on, so that one really stands out at me. So, could you explain a little more about what the intent is with that one?

- [Dr. Tcheng] Yeah, so procedures, you're right, is not really kind of a discrete concept, like all the other ones are, but it's a concept that's pretty easily modeled because we have a CPT, common procedural terminology, coding schema, so that's really what's meant by the procedures, is to have a capture of the
CPT codes. So, that's kinda the very simple answer to something to build as a specification. We did find in many of the registries, especially, obviously the ones that are procedurally oriented, they wanted to know not only what the procedure was, but what all the other procedures were that were being performed in the context of a procedure, so that's what's meant by that.

- [Elise] And I have another question. So, for the medications, and this is also something that we were sort of talking about in our project is that, so the registry, you're talking about specialty society registries. I want to point out because there's 100 major registries, there's actually over three thousand in the registry of patient registries and we think that's just the tip of the iceberg because it's voluntary registration. So, there's a lot more registries for a lot more uses than what you looked at. But, looking in the context of professional society registries, so are you talking about medications that that provider prescribed for that patient or all medications that the patient is on and how are you going to get that data?

- [Dr. Tcheng] Yeah, so in the hospital, and, I would say, the majority of the registries that we have are hospital-based registries it's pretty straightforward. It's what the patient actually received. The bigger problem that you're describing is a very real problem. We're not tackling that. We just want a way to register for the purpose. If you have a registry and you want to capture a piece of information about a medication, we're not discriminating or we're not trying to determine whether it's prescribed, whether it's taken out and stuck in, it's just whatever you would document in the registry. So, I'm gonna, I guess you could call, plead ignorance here in terms of the realities of all the complexities of thinking about registration and medication prescription, medication filling, taking it, et cetera, et cetera. Think of it, instead, from the standpoint of I got a case report form here, I'm the nurse clinician or whoever it might be, filling out the registry person. How would I fill out this information and then where does that data go? So, that's really the context. It is an imperfect world, okay?

- [Elise] But, I mean, if you show up at the hospital, they'll ask you what medication you're on.

- [Dr. Tcheng] Yeah, I don't care about that. This is registries. The focus here is what we, as the registry community, can accomplish, okay? We're not boiling the ocean here. We're only boiling a few lakes, okay? Yeah, yeah.

- [Female Audience Member] But, I mean in as much as for the registry medication to reflect membership on medication lists, I think there is clear relationship and workflow wise it's a consideration regardless of what the primary end point is regards medication. My comment is about smoking status. I presume that we're not locking ourselves into representing some of these things in the way that they've previously been represented. For example, there's a lot of ways of capturing information about smoking, right? And the members of the common clinical data that shows this as a value set, one might argue or not, the most clinically useful or the only clinically useful way of representing that. So can you say
- [Dr. Tcheng] Yeah, so I'm going to answer your question about work, all right? Part of this exercise is trying to bring horses to water. The horses being all the registry owners, and the water unfortunately, right now existing in little puddles here and there with a lack of clarity as to whether or not this puddle's safe to drink from, whether this lake's safe to drink from, et cetera. So, part of the exercise is to create a safe place, if you will, a nice lake where it's easy to consume the information. Part of the exercise today is to convince you guys that this actually needs to be done and to encourage you to go back to your leadership and say, "Okay, these are the resources "that's gonna be required to do it." Having said that, we're not draining the swamp. We're not removing all of the other lakes. So, the approach that we're suggesting here is, I would call a minimum spec, not the actual spec. You're absolutely right about smoking status. I almost took smoking status off of here because it is so bad. I really don't like it. Nonetheless, it is what it is. It's a specification from the USCDI, et cetera. So, at least as part of the exercise, we're gonna go ahead and do this. Now, every registry owner is going to have to ask a question, is this the framework that we want to use or do we like our representation of race where if it's two mixed races, we put two things, if it's more than one, if you're more than one race, then you might have a checkbox with more than one race. So, you would have the foundational five for race and the optionality of doing other things. Not the ideal state, but at least those first five things would be there. In terms of smoking, I don't see and, Ganesan, I can ask you, since you're from the College, from the American College of Cardiology, I don't see this changing because what we've done, we do put pack years of smoking, et cetera, et cetera. We've correlated the number of pack years of smoking with that outcome and that is actually really relevant to risk modeling predictions, et cetera, et cetera. Not the way that, and I'll show you what the way that the USCDI is actually framing cigarette smoking, but, so hopefully I've answered your question, at least indirectly.

- [Female Audience Member] You did and I have a follow-up suggestion, which is that if our goal is to create exchange data elements, then we should make the case to members that if they want a different representation for their own purposes, then they must provide to the group a mapping that goes to this exchange set, and they must make that available to people out in the field who are gonna try to implement it because that creates a pathway for the system to first normalize things, to the exchange format and then if they want to map it backwards to a separate format, they can and I think what the registries will find is the feedback they get is, "Well, we'll just use "the thing that's already there for everyone else." So, maybe a backdoor pathway to help maybe encourage people to conform.

- [Dr. Tcheng] Yeah, Stan, you've thought about this a lot in terms of trying to bring the clinical community together as opposed to odds and ends and budget in terms of specification, so if I could ask you to comment.
- [Stan] Well, I've, a couple of thoughts come to mind. One is, as you've indicated via model, there's incredible variation, even if you say I'm using FHIR and I'm using LOINC codes for lab data, you've got incredible variability still and so, that's an area where we're working and trying to work with experts and I'll be interested in what work has been done with medications because the opportunity there is also very, is via model. I mean, people, you slipped it in in an earlier slide and said we actually want to change the paradigm from saying, "You can do it however you want "in your system and we only care about "the exchange," to saying, "We want you to do it "in your system the same way," and there's incredible power in that. I'm surprised that we haven't had more discussion about that point because the other thing that I'd say is yeah, you're focused on registries and that's great, but, I mean, go back to some of the slides that Julia had and others and you say, okay, now I'm picky about Intermountain Healthcare, I'm not thinking about registries, I'm picky about Intermountain Healthcare and participating in registries and we have to provide data to public health organizations, we have to participate in research activities that are going on. CMS wants us to send patient data in support of claims and if we don't fundamentally model data in the systems the way it needs to go to those organizations, then, yeah, we sort of solve the problem for registries and I've got 20 other people I need to talk to and I'm still doing translations for each one of those.

- [Dr. Tcheng] Right.

- [Stan] And so, this is absolutely worthwhile and can be a model for that, but that's what you're trying to do, actually, is, yeah, let's do the registries, but then let's include even other people to say how do we standardize this so that it meets everybody's needs? And you don't have resistance and a translation every time you want to use that in a little different context.

- [Dr. Tcheng] Right. So, the good news is there is a much larger audience than just the registry community for this work. We have John Laschinger and Behnaz Minaei from FDA. Their work with the Coordinated Registries Network, The Women's Health Technology Group, there are a whole bunch of folks who are out there who want to consume this at FDA. At the DCRI, I've chatted with my informatics bosses and they're already committed to building this within all the clinical file structures of the DCRI once we have it done at the center. So, the concept is to build critical mass. We convene the registry community as a group who is logically interested, but the good news, reflecting what you're saying, is that as we've exposed this, as we socialize this, everybody says, "Yeah, when can we have this?" "We need you to move forward as fast as possible." Yes.

- [Female Audience Member] So, I come from the clinical side, so I agree with several of the comments and how this is integrated at a practice level since the data that is going to be used is coming from that atmosphere. So, I'm looking at the different social statuses, is kind of what is always the trigger and in the practice side, often what drives, unfortunately, what drives a lot of the workflows in the way that clinicians are capturing data is through regulatory reporting. And so, when, I know one of the things on the practice side that they struggle with is 'smoking status' is now 'tobacco use' and 'substance use' now
includes alcohol. So, I think thought needs to be just very prudent in how some of these elements are labeled because you could look at this and say, well, that's granular, that the smoking status actually is a subset of, say, substance use or tobacco use and so, I am concerned a little bit of how this gets integrated into the practice side when you're collecting that data and does it become an additional burden to clinicians, as this is just another driver for them while they're trying to also meet regulatory reporting and things like that.

- [Dr. Tcheng] I agree completely, yeah. Part of, and John Windle, who's over in the corner there, John and I have been working on his grant. He has an R01 grant to identify the EHR constructs of the future and a huge component of that is data persistence. So, you don't keep capturing the same pieces of data over and over again. Very sensitive to labels, et cetera, and we'll see how this rolls out here in the next few slides. Seth, did you have your hand?

- [Dr. Tcheng] I'm sorry, Seth.

- [Male Audience Member] No worries. Is there a reason to include so many, even though this is a pretty condensed list, so many data elements in the first pass? Because you're solving the same problem multiple times for things like sex, height, weight, where it's a single value. There are deeper problems, like procedures, medications that are more complex. If you narrowed the list down to three or four that represented different conceptual levels from something like sex to something like infection you could model those conceptual levels and then apply that model to other similar elements. I'm just afraid that even with this large list, with this much discordance, you're biting off years of work to get to models for all of these when you could focus your attention and work with groups that are like-minded on more complex concepts.

- [Dr. Tcheng] Yeah, so we will be finished with this September 30th. Very ambitious, okay? What we wanted to do was create something that we could hand off to registry owners to say, "Give this to your database developers "to implement." We did not actually want, we wanted an easy button, we did not want a work product that said, here is the approach that you take for doing straightforward data elements and we're only gonna give you one. We actually wanted to give them all of them and so, again, getting back to one of my first slides about audacity. This is audacious. We're trying to tackle a lot of stuff. I have a great team, that's the good news. We've been working very hard on this and bringing all this together. But, I didn't think it would be that useful. It'd be more of a white paper if we just hand it off to the registry owners. This is the model for doing this as opposed to saying, do you want to have date of birth represented the same way as everybody else is going to? This is the way you do it. And so, I think we can actually accomplish this by the end of September. That's what the plan is, yeah.

- [Seth] So, just one follow up on a few comments from Elise and then just some context briefly. So, this is not really a data structure in a sense, so there's no need, there's no need, excuse me, there's no need
to use all these data elements, they're just, if you're going to use these data elements, sorry, if you're going to use these data elements, then these are suggested ways to semantically harmonize, so let's not think of this as a case report, it's just a set of data. The idea is that this is a catalyst. These data are used in many kinds of systems. We focused on the society registries, which, at least I know, are just a small subset because we feel like together, the society registries are collecting the majority of clinical concepts that are important in medicine, so, whereas, the thousands of other small research registries are very narrowly focused on particular research questions. And so, the idea is we learn from this, we use this, try to get it adopted as much as possible. A lot of these concepts have already been modeled, so it's not gonna be years of work, but we do need to use this, but just wanted to make those comments. Thank you.

- [Dr. Tcheng] All right, go ahead.

- [Female Audience Member] I just wanted to add a little something. You know it's been, what, 50, 60 years since Loving versus The State of Virginia allowed interracial marriages or approached them to get rid of the law against interracial marriages. And so, there's a cohort of mixed race citizens of the United States I would say people accessing healthcare. What if it might be better without trying to introduce a lot of new complicated terminology and just say check all that apply because otherwise, If I am seven eighths African American and one eighth something else, do I have to go by the Passe Blanc rule of New Orleans and call myself black or can I just say check all that apply? And let's see if I can get to some of these which are going to become mixed values if you're looking to the future. But, you're not adding anything else, You're just...

- Right.

- [Female Audience Member] letting it be or what it means.

- [Dr. Tcheng] You're actually anticipating the key homework assignment, which is we're going to represent the information here straight up. What we find in the existing ontologies, the representations, et cetera, et cetera. Return these to you and say, what would you, do you agree or what would you do differently?

- [Audience Member] Right.

- [Dr. Tcheng] This is actually one of the key ones. I'm very much with you, that makes sense to me.
- [Female Audience Member] And I'm okay with sex at birth. I would say that the number of cases of testicular feminization it's particularly rare that that person is considered female. But, I like the idea of bringing substance use together as many different substances and in that case, hormone use by people who are affecting a transgender phenotype could come under substance use.

- [Dr. Tcheng] Okay, good.

- [Julia] I think part of the value of bringing together straightforward groups like this is the opportunity to educate and to create and apply best practices. When you look at some of those lists that we found in the previous two presentations, you see a lot of bad practices, right? You see coding terms like "the information's unknown" into the value set itself for the term instead of allowing the same standards to represent the information for the data elements unknown, right? Or creating local codes when codes already exist and recoding something that's already coded. So, we can help to reach out to organizations that own this content and help point out, maybe this doesn't follow best practices and let's learn about what best practices are and work to improve the content that we have.

- [Dr. Tcheng] All right, so I've got two more blocks of things that I want to get through. Did you have a question or a comment?

- [Audience Member] I just have to make this comment. I'm belaboring the obvious, but Stan brought it up and that is, you said interoperability of what to begin with? And I'm saying interoperability between what? And I'm not so much concerned about interoperability among registries as I am between registries and electronic health records because of the burden that Stan talks about and it's growing up all the time, so from the standpoint, and this is a burning platform, registries need to get to this to reduce the burden on hospitals reporting data to them because otherwise hospitals are going to start looking for alternative ways of collecting the data rather than paying for the privilege of having the burden of reporting. So, I think that these can be a little bit of a burning platform for this.

- [Dr. Tcheng] Yeah, I couldn't agree more. I think that there is a bit of a golden goose here for professional societies who built these registries and a bit of a, if you will, cash cow for those registries, but that approach is short to be here because there are gonna need to be better, more efficient ways to capture the information and that's part of what we're driving towards here. All right, let me move on. You saw this before. This is the metadata slide that Becky showed and I don't know if you had enough coffee at that point in time to really pay attention to it, but this is actually the next concept that I think is really important and that is if you're building a database, what are the things that you, as the programmer, as the database developer, need in order to build the database in a way that is interoperable. And so, we've worked on these metadata back and forth trying to figure out what is the minimum, a parsimonious set of metadata? I'm not talking about the formal ISO 111-79 - 3 dash three modeling that is the way the informatics community likes to think about metadata,
associate it with a concept. Instead, I'm asking you to put that aside for a moment and think about it from the standpoint of that database programmer who's in your basement, and you're asking them to build concepts. I'm also asking you, at the same time, to think about the data dictionary that's typically derived from that work. You have a concept and you want to present it to humans, the clinic, clinicians, typically, so that they understand what it is you're trying to capture for that particular concept in the database, right? So, these were the key CDE metadata that we've come up with. The clinical concept label, which is really the human prompt. So, if you have a piece of paper, what would that concept be labeled as: date of birth, birthday, whatever it might be for the case report or if it's electronic, what does the data entry screen say? What's the clinical definition of that? So, you print it in your data to sharing. Another way to think about it is if you're doing it electronically, you can hover over it. What would pop up as the definition for that particular concept. Where there is a value set, what would those allowed values be at the screen level, okay? What's the dropdown list look like? At the database programming level, what's the label for that field? So, is it sex? Whatever it might be. What's the data type, that format? Is it a character? Is it a date? Is it a float? Is it an integer? Et cetera, et cetera. Are there business rules that need to be applied to those database elements? A range check, so you have to be older than zero and younger than a hundred. Whatever that range check might be. The database allowed values is a little bit different of a concept than the clinical allowed values. Clinical allowed values is your list of the checkbox things or of radio buttons or of the electronic dropdown list. The database allowed values could be exactly the same or they could be a code scheme or any kind of variation thereof. Allowed values definitions, usually the allowed values are pretty self explanatory, but sometimes they're not, so whether they're not, especially at the clinic level, you need to have definitions of those allowed values. And then, the key reference ontology. This is the original, the source, if you will, for the ontology for that concept and then when we have a value set, also, the bindings for the values for that value set. And I'll show you what I mean here in a few moments. We're paying, purposefully, a lot of attention to what FHIR is saying. I think FHIR's done, Graham Grieve and his team, has done a great job collating a lot of this information and putting it in a place that has reduced dramatically the amount of work that we have to do. But, we said that in order to make this the most useful for a database programmer, they need to be able to go to a URL, click on a link inside the instructions and go straight to where it is that they can find the information for themselves. And so, we're including both the FHIR profile and where there's a resource. Some of these things have value set references, as well. And then, the last concept is sources, reference and notes. Now, here's the exercise and this is always the hardest part of the exercise. What's missing from this list that you would think that you would actually need, especially as a database programmer, but also put on my clinicians hat, my doctors hat and say what else do I want to see at the doctor's level, permissions level if you can imagine taking some of this information and putting it into a data dictionary, whatever form that might be. Is this too much? Is this enough? Is this just enough and what's missing? So, that's where I'm going to stop here for a moment unless you want me to show you an example of these. I can do the slide example, too. All right, so, I've got three examples. Let's do that. And then we'll come back to the questions, okay? So, here's sex. I think most of you, hopefully all of you, were on the Webinar that we had before. We made a big deal about the difference between gender and sex. If you don't know what the difference between gender and sex is, come up and ask me afterwards. But, we're focused on birth sex, sex as described at the, your biological sex, if you will, mainly because of the association between your chromosomes and risk factor predictions for presentation of disease, et cetera. Okay? So, the clinical concept label, very simple one, what should be on the screen or what should be on the CRF? We're suggesting sex. But, if
you’ll notice, I also put in brackets there, other ways to express this. This not absolute. You can do it any way you want, birth sex, or sex parentheses, birth sex, something along those lines, but to get across the idea that what is being requested is ‘what is your sex at birth?’ The clinical definition is a mashup of things that I found, actually, not in medical journals, but of all places, Wikipedia, plus a couple other places, but what I’ve ended up with is the person’s biological sex assigned at birth, not to be confused with the social construct of gender. So, in as few words as possible, a very explicit definition, which, unfortunately, I had to make up myself, so, it is original, but it obviously is derived from a lot of other things. Clinical allowed values, what would you see on your dropdown list? What would you see on the screen? On a piece of paper? F, M, UNK, or female, male and unknown, again, variations of that. And then, we get to the specificity of what should be at the database. So, that the database field label that we’re suggesting is sex, and you’ll notice that it’s all caps. There are some computer systems that really don’t handle upper case, lower case very well, so the convention is that we’re going to just do it in upper case. The field type is a value set and because UNK has three characters, they have to be able to capture three characters There aren’t any field business rules, there’s no range or edit checks here. The database allowed values, this now gets to what the database programmer needs to present or needs to capture in the database, F, M, and UNK. There’s actually a reason for that and that’s because of the way the interoperability bindings were. That’s what the FHIR resource and Vista page finding, so that if it’s pulling a piece of information out, it doesn’t need to interpret those concepts. Allowed values, female, male, and then unknown. Unknown has words associated with it, so define the allowed value, female and male do not. I could go on and say female is an XX chromosome, male is XY, but I didn’t think that was necessary. There is an ontology concept that’s listed there for LOINC, concept, sex assigned at birth. Then the bindings, so LA three dash six in LOINC is equal to F, that’s female. LA two dash eight is male. LA four, four, eight nine dash six. I’m not sure why it took them so long to figure out that they needed to handle unknown, but that’s what it is, okay? And then, I mentioned before, the FHIR concepts that express this are URLs and then this is designed so that you can click on it and go to the FHIR profile about, I know it’s too small for you to read, but this is a FHIR profile from a US-Core patient where birth sex is handled, okay? And then, and then, the resource itself where, go down here. Let’s see. This is an extension. Wrong. At any rate, it’s there somewhere. Male, female. Where was it? Okay, so F, M and unknown and then the code system bindings, et cetera. What I did not put on there is the OID. For those of you favorite OIDs, I cannot read OIDs. I obviously can read the numbers, but I have no idea what they mean, all right? Is it a LOINC, is it PHINVAD what is the, SNOMED, what’s the original source of that, so I actually avoided putting in OIDs because, clinically, and I would think from a database programmer level, you’re not going to remember all the OIDs from all of these, so that’s why it’s done this way. And then, the only note that I have about this is USCDI, and you can imagine this becomes a spreadsheet, all right? I’ve only extracted the key pieces of information from the spreadsheet. There’s actually more detail, but a spreadsheet would be even harder to read. Harder to read. All right, so let me go back here and stop here again to ask the question. Is this the right list for what you need as a registry owner to understand these concepts, to make sure that they are either consonant with your representation or they are part of your representation and enough to hand off to the database programmer without being, if you will, too much. Question, yeah.
[Male Audience Member] So, I think there’s a couple of concepts. One is cardinality, might be important to include. The other thing is, at least from your example, you said that you were getting answers, but not necessarily displaying the question asked.

[Dr. Tcheng] Yeah, so the question asked is actually the clinical concept label. That’s the question, sex.

[Audience Member] Okay, and then using the LOINC code or assigned code with that, so there’s specific codes for that and I don’t see your reference ontology for the question, you just have it for your answer.

[Dr. Tcheng] Yeah, so, the concept is here. This is the question right here, LOINC LL three three two four dash two, the question actually is sex assigned at birth.

[Male Audience Member] Okay.

[Dr. Tcheng] Yeah. So, you’re absolutely right. You have to have representation of the question and you have to have representation of the answer in the value set, as much as they exist. That’s what’s meant by that.

[Male Audience Member] Two things. One, you mentioned that you have deliberately dropped the OID and I don’t like it either, but how do you ensure uniqueness of your code when you have your value set that could come from different systems.

[Dr. Tcheng] Yeah, so, the specificity here, for example, in this example, the value set, the bindings, it comes from LOINC. LOINC has, it’s an OID, so it’s there. That’s the one that we’re recommending. We’re not saying you ought to use the SNOMED representation or anything else, yeah.

[Male Audience Member] So, you would not have value sets that come from LOINC as well as SNOMED at the same time? Is that the assumption?

[Dr. Tcheng] No, yes.
- [Audience Member] Okay. The other thing I wanted to point out is the context, temporal context, our registries capture that information. Most of the time it's metadata driven, and also the work flow context in the data has been documented.

- [Dr. Tcheng] Right, so one of the things, and John, this will resonate with you in terms of the work that we've been doing with The Women's Health Technology Group, as well as others, CRN, Coordinated Registry Networks at the FDA, is they come up with a list of metadata, so it's really kind of interesting how all the efforts that convene still have the same set of problems. But, you're absolutely right. One of the things that were in the Women’s Health Technology CRN that's not on here is the very bottom is the classification or context. Is it a demographic? Is it procedure related? Et cetera, et cetera. So, I've purposely not talked about context at all. It is a key issue. For example, if you're performing a coronary intervention, and you run into all the other procedures that are performed in the context of that coronary intervention, then the context becomes important. Sex doesn't. Birth date doesn't. But, medications do, labs do, et cetera, et cetera. So, you're absolutely right. That's one of the things that's not on here. Should we list all the context that this will be applicable to, so I'm happy to entertain that concept, too.

- [Male Audience Member] What about the temporal context?

- [Dr. Tcheng] Yeah, so, temporal context I think falls within that framework as well.

- [Female Audience Member] So, how does, how does this merge with SDC? Just one way of including unit? How does this merge with SDC? So, originally when FHIR started, FHIR adopted the SDC mode of transfer with a form file that contained a lot of the metadata that you list here and that FHIR slash SDC one-to-one marriage seemed to disappear around 2017. So, where does FDC fit in with this?

- [Dr. Tcheng] Yeah, so FHIR only did a little bit of work with SDC and then kind of, I hate to say it, but gave up on it as a way to try to integrate the capture of data through the process of care, so we basically, like I said, we built our own at Duke. I'm not really sure about how many legs SDC has and whether it's the way of the future, et cetera. I'd be more interested in your perspective.

- [Female Audience Member] So, the College of American Pathologists has adopted it for the electronic cancer check list and in 2019, they'll be released as SDC with their form files with them. They will still look like XML and they'll have a lot of the metadata that goes with XML, so be structured in the way the XML is structured. But, in terms of interoperability, to collect data, transfer digitized data or transfer data as you want it to be viewed, they'll have both available.
[Dr. Tcheng] Right, so I'm picking a cardiologist perspective number one, so you'll have to excuse that, but in terms of kind of the SDC, the structure of data capture concepts in terms of surfacing forms the issue that we had was that you had to capture the information and then you surfaced the form and then you brought the information forward. That just didn't make sense. We just wanted to capture the information once as part of the process of care, so that's where it fell apart in terms of it didn't reduce the extra work, so.

[Female Audience Member] So, in regards to SDC, and I'm assuming, so, it can be confusing because people talk about structured data capture, general activity, and I think you're referring to the Structured Data Capture Initiative, which is based on ISO 11179-3 (dash three) form and matter for data elements, right? So, SDC was created as an extension in FHIR and was supported through a funded opportunity out of ONC and I think there are two issues. Issue one being that the primary use case was a research based use case and so, people who tried to use it as a point of care form and matter for data found that it was too cumbersome to actually effectively integrate into a workflow. There's an interesting project that did, if people are interested, where they created mini versions of SDC style research data elements for use in an EHR point of care setting. So, that's why I'm, I think the challenge is that there's no, because of that limitation, and I don't think there's a pathway to really onboard SDC into FHIR core for the purpose of clinical data elements day-to-day and the other challenge being that the funding from ONC is no longer running, so it becomes a question of who's the owner of SDC now and will there be a future pathway of integrating SDC into the primary point of care use case in a way that makes it useful? It's very possible and actually some work was done on this to display a way of showing SDC's extension to a smaller version of, of the core data elements. There has to be commitment to create and maintain that alignment and, to this point, there wasn't that commitment and so, there's a question of whether or not that will continue to be a viable standard.

[Dr. Tcheng] So, Becky, let me do a time check with you. Do I have until now or until 12:15?

[Dr. Tcheng] Oh, okay. Okay. All right. Very good. Let me do this then. Since we talked about smoking status, I just wanted to show this to you, that we've also modeled this based upon the existing standard or, I should say, model. It's not the right phrase. I'm showing you a representation of it, so the concept label smoking status, alternative smoker, tobacco use, tobacco smoker, none of those are original to us, by the way, these are all ways that they have been presented in various resources that we identified. The clinical definition of current smoking status of an individual. The lab values, and Julia was mentioning before, this is where it really doesn't work very well. The allowed values are derived straight from the existing value set from the CDC, and that is current heavy tobacco smoker, current light tobacco smoker, current some day smoker, I'm not sure that means I'm gonna smoke someday or what that means. Former smoker, never smoker, smoker current status unknown, current everyday smoker, I'm not sure, again, how that differentiates from the other ones, unknown if ever smoked. The database field level we're suggesting is smoking underscore status. The value set or the database field type for the
value set, and you see the number of characters that we need to include in order to accommodate the
text. The database allowed values, here's where the values actually become exactly the same as the
clinical allowed values. The allowed value definitions per the CDC. I didn't reproduce them, there are
actually some fairly specific definitions according to the CDC. The concept, itself, of smoking is
represented there in LOINC. The value set bindings are from the CDC, which actually come from
SNOMED, all right? But, the first one, which is current heavy tobacco smoker is four two eight zero
seven et cetera, et cetera. The value set is here and, again, just to show you, look down here at the
codes, okay? The codes and what the display should show. And then, this value set contains a concept,
that actually comes, like I said, or is instantiated in the USCDI. If I can find it.

- [Male Audience Member] Are you going to get to the level of saying what you mean by previous
smoker? How long they have to have been not smoking to be a previous smoker?

- [Dr. Tcheng] Yeah, so I didn't put it on this slide just because there's not enough space, but the CDC has
all those definitions in the public health information network value or something or other, which I can't
get to quickly, but, it's one of the links that are here, all right? All right, so, I'm gonna close then back to
this query, and that's the question is, are these, and what we're proposing is on the left-hand side, are
these the right ones? Should we include cardinality? That was a suggestion. Should we include at least a
description of context, including temporal context? Think about it, again, from the standpoint of a
database developer. A database developer, that's the primary context here. So, I think cardinality is
important. I agree with that. Any other comments? Anita, you've been thinking about this a lot for the
CCRF project.

- [Anita] I think about version. I think there's a comment, too.

- [Dr. Tcheng] Version control, okay.

- [Anita] Thinking about the future and if you have changes to it, you probably want to consider the
version of the data element.

- [Dr. Tcheng] Yeah, so for right now, the version's going to be the September 30th version because
that's the end of the grant, that's the end of the project. But, hopefully this will have some legs and we'll
keep going. Other comments? Questions? Thoughts? All right. So, again, what you're going to be
receiving towards the end of September is this. It's going to be all the data elements that we've
described, all, I think there are 15. Data elements is used, the term's used loosely. Look at the question.
We need you to take a look at every line and say yes I agree, no, I don't. The ones that will most likely be
subject for discussion I think are smoking status and race, based upon the things we've chatted about
here already. Okay.
- [Julia] And, Jimmy, it would be really cool if there was a training kind of style webinar where we took a bunch of data elements from various candidate sources and walked through how would you map out the pieces of this because I think it's hard for a registry to think about this abstractly. It's much easier to sort of see it in motion.

- [Dr. Tcheng] Yeah, so, we have one more webinar towards the end of September and we can do exactly that. To illustrate how it would be actually built in the database. All right. I'm gonna hand this over to Tom. Oh, one more question, yes.

- [Male Audience Member] Yeah, I don't know if this would fall, ah, thank you, I don't know if this is outside of the scope of this, but the only thing I can tell that's not being represented is the place where the data is being collected. Is it being collected by clinician, point of care, directly from patient bureau, that kind of fact, but I don't know if you guys would, if that's something you would normally include here.

- [Dr. Tcheng] Yeah, so I think that's more of an operational thing. So, the question really is, what should, again, think you're a database programmer in the basement of an office building and you're being asked to build a database, but to capture the information from a registry or for a registry, how would you represent that data in the registry? So, the operational side of who does what, when, where and how this is being collected, I'm not so certain that that falls within this. So, we'll consider it, okay? All right. Tom, if I can ask you to come on up here and chat with us about your work in terms of thinking about medications and medication models. And, Becky, time check. How long do we have for Tom?

- [Becky] Tom has until 12:30.

- [Tom] So, I'm here to talk about the medication model that we have been working on. From this, I specifically reviewed 36 different registries, of those, 28 had questions directly related to medications and there were, as you can imagine, a large variety of specific questions, but we had four general categories of medication questions. The first type of question was an open-ended provide a list of medications. This could be on MedRx or similar format. So, a clear example of that is it lists all medications the patient is currently taking. The second category, and there were 12 registries that asked for them. The second type was the specific domain-pertinent medications. Is the patient on a TOBI Podhaler? That sort of question. Very specific domain that has very little crossover between registries. The third type was medications in a drug class. This was the most common. Ace inhibitors, anticoagulation. There was no majority class of drugs that was asked in registries. And then, the final category was medications that are related to procedures and from these four general categories that we identified that all have done a preponderance to be worthy of study by this research project, we found
16 specific types of questions that can define these four. Those 16 all have existing sources that were referenced by the health information technology standard panel that was an AHRQ project from a few years ago that identified specific sources. So, fill status, for instance, has the patient been, do they specifically have an order that's been prescribed? And that's HL7. What is the indication, there are a few different standards for some of these, so, ICD-9, ICD-10, SNOMED CT. Product form, that would be the physical form that the product is presented to the individual. Dose, route, type of medication, that's the marketing. Site, the anatomic site. There was no specific standard identified but I think SNOMED CT would probably be sufficient for that. The brand name. So, what is the brand name of the drug? That was RxNorm. RxNorm is an NOM standard that has a specific identifier. So, instead of... in the packaged product, which is down a little bit lower that's an NDC. So, that would be a pharmaceutical packaged product, so is X drug, 30 pills with X milligrams from this specific manufacturer. That's a more pharmaceutical question. And we didn't find any, in the registries, there weren't any questions that were that specific to that level, that degree of specificity, but we want to know that there is a standard for that that we can utilize. The big, one of the big ones is drug class and the identified standards were RxClass, which is a collection of standards and then NDF-RT, but NDF-RT is actually going out of style on January 2nd, I believe, and it's being replaced with Med-RT, And I think probably the most... and then the three that we identified that were outside of the original AHRQ model were the prescriber, the effective date and the duration because there were a lot of questions about "within the last 30 days, have patients taken X?" Now, who adheres to these standards? In drug class, of the 18 registries that asked about drug class, a total of zero. No one adhered to the established standards, or at least no one declared that they had. The clinical drug name, which is a combination of the list all medications and specific medications, RxNorm or LOINC or similar were only used in nine of 21. So, from here we can see that there was just not a lot, in medication, there's not a lot of adherence to established standards. Not enough for any preponderance according to our study, except for effective date, but that was more the standard of date, time is so well ingrained in all of us that we all just code it that way. Now, from this, I'd say the big conclusion of this is medication data is necessary for registries to be interoperable from vendor to the registry. And there's a lack of adherence to these standards, or if they're being built with standards in mind, there's a lack of communication of these standards in data dictionaries. And so, the need to communicate these data standards to local builders, I think, is one of the big take homes from this and in our data dictionary building to make sure that people know what you are trying to define and to utilize existing standards. I think we can go into a lot more detail in specific areas, but the key of it is, at start, define the variables and then in the next month or so, we'll come out with a paper where we define what our recommendations are for the guidelines. And then, from that, it's imperative in the data that we're sharing to straight up just say, "This is what we use and this is what you should use." "These are the RxNorms, CUIs that are acceptable for this." "This is what an Ace inhibitor is," cause a lot of times we're just putting it on clinicians saying this is, "Well, were they on anticoagulants?" rather than going, "Were they on anticoagulants?" "Here are the lists of anticoagulants." It also prevents us from automating. So, pulling that data out of EHRs and then pushing them into registries automatically. We create another human step by not adhering to these standards. So, are there any questions about medications? I know I went pretty quickly, but any thoughts or questions?

- [Female Audience Member] Hi, this kind of piggybacks on Dr. Rich's comment earlier about those of use who are using the technology in our registry to pull the data in from EHR directly. I think we're
finding a lot of the detail of this is in the physician note, or in actual text fields and not in standard data fields.

- [Tom] I would whole-heartedly agree with that, especially in medications. Some vendors have multiple places where you could find medications and there's lack of consistency internally within their system, let alone in communicating that without set systems. I would say that's a real problem and is a key finding that we need to communicate to vendors.

- [Female Audience Member] Yeah, I just wanted to add to that because even within the structured data fields within EHRs, it's difficult. So, I did a lot of mappings to pull medications for ECQM and found, depending on the vendor, you would think that our accounts would automatically be connected to a medication and they could have been in the order field, but then in any other field, there would be no RxNorm. So, to be able to pull the medications was such a challenge. Medications is very, very complex and you wouldn't think that because it's very straightforward, right? The medication and the RXNorm, but the way that they're structured in EHRs, they're so complex.

- [Tom] I would agree with that. I think it's a part of the problem with getting a good, clear definition of what medications are. Different stakeholders have different needs. So, pharmacy notes have to be different from clinician base and that's where the big NDC or RxNorms what happens. It is imperative for pharmacies to use NDC codes in case of drug recall, in case of specific generic issues, as we were recently saying, whereas clinicians, it's more just note what the patient's taking and then possibly be, ideally, we would be able to roll that up into a class.

- [Female Audience Member] Yeah, I think that it keeps coming back to the need for a model that reflects an appropriate workflow because the NDC isn't generated out of thin air, it's generated from the order, which should have an RxNorm that tags it. So, if we create a framework that allows the metadata to build and add to the longitudinal care activities and there's a billion more potential use cases that that could support, but if we don't provide leadership to create that and then try to use levers, like ONC, to push that through to systems, systems can still do whatever they want. So, very frustrating, and I'm sure Jill can attest.

- [Tom] Yeah, I mean, in terms of drug mapping, there are some NDCs that do not have RXNorm CUIs associated with them that we found, but I have not done any literature that proves the other way, so every NDC code does not have it. Yeah, there's not 100% mapping, even between those two critical data sources, let alone vendor specific drug codes that they might have put in their system before they even knew what RxNorm was.
[Dr. Tcheng] So, Tom, you've got more than a dozen different variables here. I presume if you look at the information that the registries are trying to capture, none of them are asking all 16 or 17 and I think that's what you showed before. But, if you were to try and encapsulate what the model should be for registries, how many data elements are we talking about relative to every drug and then is there kind of a model that you're thinking about that really makes the most sense right now?

[Tom] So, from the four types of medication questions, the 16 data elements are the core of those. There are some that are just theoretical that I used, so we could probably pretty easily bump it down to eight. I don't think a lot of people need form or vehicle in their registries. I don't think we're at that level of specificity. I don't even want to be at that level of specificity. But, that was just coded in the original ARC code, so yeah, we would be able to knock this down pretty quickly.

[Dr. Tcheng] How many are you thinking?

[Tom] I mean, just based off of our information, I would say about five, six that are core, drug class, drug name, dose, the effective date and duration, I know duration has fewer, but there are enough conceivable questions that have to do with pre-procedure status and post-procedure that I feel like both effective date and duration would be important.

[Female Audience Member] So, are you including units, also in the dose, or where are units located?

[Tom] Yeah, so for this, yeah, so, in dose we would use the NCI, LOINC, or HL7 v3. So, yeah, the units would be inherent in the dose.

[Female Audience Member] Okay.

[Tom] I believe. I'm sorry, it's a very small font. But yes, from there, it's part of the plan. I didn't want to get too specific in terms of recommendations because, again, the ideal goal that we all use the same language between all 40 registries is great, but if we only get to the point where we just communicate what terminologies we do use, and every, all 40 registries in their data dictionary go, "I'm using the LOINC standards," "I'm using the HL7," that would be such a big step forward, especially for developers just to know what they're doing and know why they're doing it that I think that would be, if nothing else, for medications, if we make no bigger step than that, that's a big success. If we get to the next step, that's a bigger success. Well, enjoy lunch everyone.
- [Anqi] So, one more thing before we break for lunch. So, after lunch, I believe around 1:15 or so, we're going to be splitting up into four different work groups. As you can see, there is one on return on investment, the second one is on communications, third is technical challenges, and the fourth is on adoption strategies and how they relate to some of the policy issues, whether it's a government related. So, we are going to be not all in this room. So, there are signs outside each of the rooms for the break outs. I've also listed the floors and we'll have people that can help direct everyone to the right place. We're letting you guys choose which work group you want to join. However, if there are no more seats in a particular room, if you could please select a different group, that would be much appreciated. Did I cover everything, Becky?

- [Becky] I think so. Lunch is out in the foyer where the breakfast was set up, so you have until 1:15. So, enjoy lunch.