Wendy Funk: Hi, everybody. Good afternoon or good morning, depending on where you are. My name is Wendy Funk; I work at Kennell and Associates, and we work under contract supporting the DaVINCI Project. And so, I’m here to talk to you about the DoD data helping to bring out how data are collected in the DoD systems, so we'll talk about CHCS, AHLTA, and a little bit of MHS GENESIS. We'll talk about how our EHRs interact with the Defense Enrollment Eligibility Reporting System, this is where we get our, I guess it makes sense, eligible enrollment. And then, what kind of impact, how this data gets put out under the OMOP model.

The webinar that I’m giving today is the second in a series of four where we're focusing on the DoD data, hoping that if folks can attend all four webinars that they'll be able to get a really good understanding of the data that we make available. So, this is Webinar 2.

So, the objectives for this session would be that attendees could describe the sources, the processing flow, and the completion lag of DoD EHR data, and DEERS data and where they interact, provide an overview of main direct care files, describe limitations of inpatient professional coding and how this might impact your studies, and also describe non-clinical encounters that are in the direct care data and how that relates to OMOP.

So, just a very quick recap of the Military Health System. And for people who've seen me speak before, you may have seen a few of these slides and I apologize for that, but you don't you never know who's been to every session. And so, what is the MHS, the Military Health System? It's a network of military hospitals and clinics supplemented by programs that enable beneficiaries to seek care in the private sector in order to fulfill their healthcare needs according to access standards and to assure medical readiness.

The mission of the MHS has really evolved over time. And when I first started working with the MHS, about 85 percent of the care provided was in the direct care system, so these would be at facilities like Walter Reed, or Navy Medical Center, San Diego, or something like that, and only 15 percent of the care was provided in the private sector. But in around '94, Congress instituted new laws that require the MHS to meet access standards; if they couldn't meet those access standards, then care would get referred to the private sector. And today, private sector care represents about 52 percent of the care that MHS beneficiaries get.

So, that split between direct care and purchased care has a really big impact on the kinds of data that you have available; because with direct care data, it's EHR and you get a lot more clinical data and things like that; and with claims data--purchased care, it's just claims data, so you're kind of stuck with what administrative data is available on a claim. So, the Military Health System has a robust system between both direct and purchase care, but it really is quite fluid the relationship between the two over time.

Here's just a quick snapshot: there are 9.4 million eligible beneficiaries, about half of those beneficiaries enroll in our HMO called TRICARE Prime--a little more than half, actually; there's a little more than 50 hospitals, some of them are GME facilities, some of them are quite small--we have one that has an average daily census of about one,  it's in a very remote location; there are more than 500 medical clinics ranging from Troop Medical Clinics to clinics with full-service same-day surgeries and things like that; 300 dental clinics--our dental clinics really almost entirely provide care to active-duty inpatient admissions, almost a quarter of a million, at military hospitals, but you can see the balance I was talking about where there's three times that in the private sector hospitals. That's because a lot of our locations, our bases, don't have inpatient facilities; they just have outpatient.

For office visits, there are 41 million office visits in military hospitals and clinics; and in the private sector, there are 86 million office visits. For prescriptions, military pharmacies, 34 million; private sector pharmacies, 55 million. So, you can sort of see, in this chart, that split I was talking to you about between direct and purchased care and notice the preponderance of purchased care.

So, the Military Health System Direct Care System. This refers to the acute care hospitals, clinics, and dental facilities operated by DoD. I’m only showing the United States on the charts that are to the right or the maps on the right, but those blue dots represent the military hospitals and clinics throughout the US. You can see lots and lots of them on the East Coast, not so much of them where there aren't as many people up there in the upper Midwest.

Most of our hospitals are small; there are only six hospitals in the whole MHS with a daily census of over a hundred, and we have scores of hospitals with daily census less than 50. Obstetrics is our most popular product line, more than half of our admissions to our hospitals are obstetrics-related. Some of our hospitals have GME; as I said, clinics can really vary and we're really into patient-centered medical homes just like the VA is.

One thing that I think is important when you do studies, is there is no cost-sharing for care at MTF; so, you don't have that barrier to care that, "Oh, do I have enough money to get treated?", and that can be particularly useful when you're trying to look at an elasticity of demand type study, you can compare the direct care to the purchase care and look at utilization where people do have cost shares versus where they don't.

There's also an established priority for care, and so not every patient is treated the same. The Military Health System has to treat an active-duty service member above, say, a Medicare-eligible person; and so that priority is important because at busy MTFs, what you'll see is that they'll treat a lot of active-duty and active-duty family members, but not a lot of retirees because they just don't have the space for that in some locations.

The chart that's showing you the lower half, the map there on the lower half, is the new market structure that the MHS implemented last yeah. So, this is a new sort of command-and-control type thing where there's a market manager and then within each market, a whole bunch of MTFs, and the market manager is responsible for all of the care in the local area.

So, the kinds of data that are available--this is coming from CHCS and from AHLTA--pretty much everything on here is available for GENESIS too--I’ll make a point to tell you if it's not available for GENESIS.

But we have acronyms in the MHS; I think I’m not going to really speak in terms of acronyms I’m just going to talk more about what the files are. So, we have an inpatient hospital file that represents institutional care, DRGs, and stuff like that. We have professional services files, so this is all the care that the doctors provide with CPT HCPCS codes; we have an appointment file and this includes, "Left without being seen," and "Canceled", "Kept", and all of that. We have referrals, so our referral file we can tell what we've referred into the MTFs--specialty care and MTFs, what kind of referrals we've written to the private sector; we have the chemistry file that has the exams and the results of the exams, and these lab files I’m talking about have the link nomenclature in them. There's microbiology exams and susceptibility files, I know you guys have been using those for some of your research.

We have some pathology files, these are text notes; radiology files are also text notes; we have the schedules, so here's where we can look at what's booked out for the next 30 days or something like that; we do a lot of performance measures on that for appointment template management.

For pharmacy, we have dispensing events; we do have clinical notes data, but we only have that data available if AHLTA is used, and AHLTA is really used for office-based care; so, when care is provided in an office we get the clinical notes.

We have an immunization file which has been very, very popular these days with COVID; and then we have Vital Signs--and the vital signs are not just vital signs; it does have the standard vital signs data, but it also has questionnaire type data like alcohol questions, tobacco questions, pain scales, things like that. So, those files come in regularly and routinely and are available to us to do research or whatever we want to with the data in the MHS.

So, we've spent the past several years taking this data and taking it out of the source table files and putting it into this OMOP file--and I know this is a really sort of messy chart, but I’ll get into some of the details of this chart in a few minutes, but I just wanted to point out the value of using the OMOP model is that if I had to write one query, let's say I wanted to look at what kind of things are happening in the emergency department. A simple query, right? But there are direct care emergency departments and they use one set of rules to create their data; and then there's claims data that has emergency department in it as well.

And so, typically, what a user would have to do is they have to be an expert in both files because the business rules are different for pulling emergency department care out of files. But with the benefit of the OMOP data, if I were looking for procedures that occur in emergency rooms, I can just go to the procedure table because that's where we dump all of the procedures regardless of where they are done. So, that's the advantage of OMOP; and with the data that we provide for DaVINCI, you can look at the data both ways. So, there's source data and there's also data mapped to the OMOP model.

So, reporting DoD care, we're going to just talk a little bit here about the source data again. So, I think one of the most important fields in Military Health System data is the treatment DMIS ID. So, "DMIS" stands for Defense Medical Information System and it was a legacy system that DoD had up to the year 2000; and when that system came out, we started this construct of a DMIS ID. These are very similar to your VA station codes; so, these are four-digit codes, there's actually a structure with these codes, so we have like a parent and a child and things like that.

This is used for reporting direct care data; and one thing that I think it's important for you to know is in our data, we still call it direct care, but it's a little bit of a strange beast; it's called the External Resource Sharing program. And what that is when we have a military provider who, they want to keep their skills up and do things that the MTF might not have the infrastructure for them to provide the services they want to do, what the military will do is enter into an arrangement with a non-military facility and they'll give the military provider attending privileges at the non-military facility.

One example of that would be at the VA. So, we do have some resource sharing agreements with veterans' administration facilities. And so, when our doctors go to deliver care, if they deliver the care in the veterans administration facility versus their own facility, they use a different DMIS ID. I’m using an example down here, if you see the table and you see this DMIS ID, 545486. It says, "External Resource Share. Air Force. Sunrise Hospital." So, what that is that's where a doctor is going to this sunrise hospital location, they're coming from an air force facility; and the what's going to happen is there's going to be a claim to cover the hospital care, the nurses, the room and board and things like that, so that's going to be purchased care. But then the provider providing the care, that shows up in all of the direct care files.

So, that's an ERSA, we call that. And I’ll show you another picture of how ERSAs look.

Anyway, the four-character code represents the facility where care is provided or where the ERSA provider is assigned; and then DMIS IDs fall within a hierarchy, so each DMIS ID has military service, a command a parent, there's a market that I showed you and then the markets can aggregate to regions.

One thing that I think that you should know is that the military service has changed starting this year; most of our facilities are under DHA authority now rather than under the authority of the military services; so, when you look at the data don't be surprised if you see a lot of things showing up as DHA, it's because the whole command and control structure changed as a result of the National Defense Authorization Act of 2018, so it's a whole different thing now.

Here's a picture of the market structure that I talked about. So, this is the tidewater area. So, we have a large--it's near Virginia Beach, we have a large facility, Navy Medical Center, Portsmouth, out there. And so, what happens is in these markets regardless of what branch of service the bases are in the market, the largest MTF in the market manages all of the facilities within the market, and then the market manager reports to a higher authority at DHA. So, it's a new construct and when you work with the DoD people, they probably will want to see data at the market level because that is how we're indeed managing things these days.

Branch of service. So, this is an important field and the reason I want to show it to you is a couple of reasons. First, if you want to identify those VA resource sharing events, you can see how you can do that if the branch of service is 1, 2, 3, or 6; that's going to tell you the facility is a VA resource-sharing facility, and it's a 1, 2, 3, or 6, it varies based on whether the agreements with Army, Air Force, Navy, or DHA--and as I noted earlier, most things are moving to DHA. You can see A, F, N, and P, and then there's civilian resource sharing too.

When I use the word "line", what I mean is something like a mash unit or a cash unit.

Another reason that branch of service is important is up until this year, we had three separate services that managed all of the facilities depending on what base they lived on, and that meant we had three separate sets of policies and sometimes that resulted in differences in the data. So, for example, the Air Force, they don't use CHCS or AHLTA for their primary immunization system, they use a system called ASIM; but the other services do and so when you try to look at immunization rates, it looks vastly different depending on the service, and that that's important sometimes to understand that. There are other instances where the policies are different; but what's happening by and large now is that everything goes under the DHA authority; the idea would be we would consolidate to one set of reporting requirements and everything would be the same. So, things are changing a little bit fluid here as we get through this process, but there are differences in data based on service especially the historical data.

Well, here's a picture of resource sharing. So, 5401 is representing a hospital in Newport, Rhode Island; it's the civilian hospital in Newport, Rhode Island. And what you can see here are four records--it's the same person--and notice that they're just sequential days. The [MCS] code, for us, is kind of like your stop codes or your... what do you call them? The treating specialties, I believe, in the inpatient environment So, this MCS code right here is repping or representing orthopedics. So, what this is doing--I actually looked at the CPT codes on here and one of them represents a surgery, but the rest of them are just rounds, follow-ups, and things like that.

So, this is direct care and I can tell it's resource sharing because of the branch of service of "R" tells me that's Navy resource sharing. When I go to look for this patient in the claims data, I have this pseudo person identifier and if you look, I can find this person and they were in the hospital from the sixth to the ninth. And if I look at the name of the hospital, that's the Newport General Hospital. So, there's a connection between the DMIS id and a connection between the NPI on the claim, and then the same person and you can kind of see the dates and you can see how these two events together represent an episode of care.

This is not the norm, but it is quite common, though, especially at smaller MTFs where small MTS might sort of do this for an entire product line, like they might decide that, "We're going to do all of our obstetrics care as civilian resource sharing." And so when you go to use the data, it's important to understand that because it doesn't look like the rest of the data. But it's not hugely common for the large MTFs; but for the smaller MTFs, it can be a really big deal.

This is the transition to the DHA authority that is good for you to understand. So, I just picked a few hospitals and I compared 2018 to 2019. And what you can see just if you look at Naval Hospital, Jacksonville is the first row; you can see that its DMIS id military service is in. But if I go down to the second chart--and this is one year later-- you can see that the DMIS ID, Military Service is "P", so you're probably wondering what does "P" stands for. If you've worked for DoD for any time, you'll know that DoD uses the word, "purple" to describe itself, and so the "P" stands for purple.

And so, what happened in 2018, Naval Hospital, Jacksonville was part of Navy Medicine East, you can see that in the command, and it was supervised by the surgeon-general of the Navy. But beginning in 2019, Naval Hospital Jacksonville is now supervised by the market manager in the area and then DHA above that. So, when you report by the branch of service, that's important.

People often wanted to understand what base was affiliated with the hospital and so we created another variable--strange name--but it's called Readiness Service, and the readiness service will give you the historical affiliation; so, you can still split things by Army, Navy, Air Force, but it's probably important for you to know that's not really how things are managed anymore.

Parents and children. So, when you talk to people about, "How many visits did we have at Walter Reed?" That definition of, "What is Walter Reed?" is nebulous; there are two approaches to looking at that question--and I think you have the same thing in your system, in the VA system. The language that we use is use the language of a Child DMIS ID or a Treatment DMIS ID; and then we use the terminology of a Parent DMIS ID. And what happens with the military is you'll have like a pretty large hospital on a lot of bases, but then you have smaller clinics surrounding the bases. I was born at Fort Hood, so I like to use Fort Hood as an example. They have Darnall Army Community Hospital on the base, they have three or four clinics; and then off the base, out in the community where a lot of the people live in Parker Heights and Killeen, and things like that, they also have facilities.

And so, all of those clinics are sort of under the command of the hospital commander at Fort Hood, and so that hospital at Fort Hood is called the Parent DMIS ID; and then all of the children are the surrounding facilities that report through that hospital. So, that's kind of the typical way to explain parents and children.

Parents and children aren't always nearby, though. Sometimes, the child facilities are far away, and so when you try to think of things from a geographic perspective, sometimes this parent-child thing causes problems. The example I’m showing you right here is Naval Hospital Lemoore. So, Naval Hospital Lemoore--it's actually Naval Health Clinic Lemoore now, it's downsized to a clinic; but at the time I took this picture, it wasn't a clinic. What you can see, it has its own DMIS ID, so that's the first row.

The second row is the Naval Branch Health Clinic at Fallon Navy Base, and the Fallon Base is actually in Las Vegas, and Lemoore is in California, so they're not horribly far but they're not even in the same state, they're not in the same geographic area; Lemoore is somewhat remote and, obviously, Las Vegas is not remote. You can also see a dental clinic under there, you can see an External Resource Sharing clinic under there, and then these operational forces, it's mostly a--it's not a treatment location, its DMIS ID is used in a different context, and I’m not going to get into that today.

But that's kind of the parent and child. So, when people ask me a question, "How many visits did Lemoore have?", I always have to clarify, "Are we talking about the big base or are we talking about the little facility?" It kind of depends on the question that people are asking.

So, here's an example of the impact of the difference, though. So, I brought up Fort Hood; the query that I’m writing here is coming from the pharmacy data, and what we're doing here is the parent DMIS ID is always 0110, which you can see from the first row, represents Fort Hood. But you can also see all of the clinics underneath: so, the Monroe Clinic, the Troop Medical Clinic 14, the Moore Clinic, et cetera, et cetera. And what this chart's doing is showing how many prescriptions those locations filled; and if you look at the percent of total, you can see it's like a 27 percent difference whether you interpret someone's question as the main hospital or the geographic area. So, just understanding which DMIS ID to use and what those mean could be important to what you want to look at the MTFs.

Capturing the location of care for DoD direct care data. So, all of our data files that we present to you are going to have DMIS IDs in them; the treatment DMIS ID is always going to represent where the care was delivered except for one exception: in the immunizations data, it's either where the care was delivered or it could be MTF recording immunization that are brought to them maybe during a primary care visit, they get a shot record and it says, "I got this immunization for my kid at school." Then the primary care doctor will actually put that into the immunizations module in AHLTA--and we can tell whether it was given by the MTF or self-reported given by somebody else.

We aren't going to talk about the TED files today, the TRICARE Encounter Data files, but those are the claims data--and claims data don't have treatment DMIS IDs in them; treatment DMIS IDs are really more meant for direct care. So, our purchased care data is going to have things like national provider IDs, or provider tax IDs or something like that to identify the location of care.

So, there's an OMOP table, and it's called the care side table. And so, those are going to be unique locations where healthcare is delivered and it's not just hospitals, it can be clinics and things like that. So, what we did is we made the primary variable for the care site and DoD data to be the treatment DMIS ID; so, that's why this concept is important to you. This table also includes the VA facilities and parts of VA facilities just like with the direct care data. So, we're going to cover also purchased care later, but there are also purchase care locations within this file.

Now, there's also a provider table. So, the providers are just individuals who provide healthcare; sometimes, people think of providers just as physicians and things like that, but this table includes more than that. It includes things like PAs, or nurses, or physical therapists; you don't have to be a doctor to be in this file. And then these providers can link to healthcare records, they can link to the person table, they can also link to the care site table so we can tell where providers work, we can tell if they operate as primary care managers, and then you can look at were they linked to visits, procedures, diagnoses, things like that. So, that's the provider table; you've got lots of information in there as you can see: how old they are, their gender, their specialty, NPI, DEA numbers, things like that.

The provider data that came in to DaVINCI from DHA is really very dirty; the DHA, depending on the data file, we use different identifiers for providers, so you're going to see a lot of different identifiers in here--National Provider ID, what we call the EDIPN, the Electronic Data Interchange Patient Number, this is a unique number assigned by DEERS; there's the CHCS provider ID--I would be very careful with that, those are not unique; those are like you might have a provider named "John Smith", well there's lots of John Smiths, and none of our CHCS are connected and so John Smith could have a different provider ID on each CHCS host. There's a tax ID here, the DEA number; we have social security numbers, there's our managed care support contractors have internal provider numbers that they use; the AHLTA uses a different provider number than CHCS does.

So, the provider data was really a mess; and really, we had to put a lot of work into making sure we understood who was who so that we could assign one provider ID to represent the same person and not have multiple providers with multiple IDs. So, we did that work in processing the DaVINCI data before that came over to you to make it easier for you guys when you're trying to use the data.

For provider specialty codes, we have three different code sets in here: we have the Direct Care Provider Specialty, and that's useful if you want to discern interns and residents from regular providers, there's separate values for those; if you look at the HIPAA taxonomy code, interns and residents will typically have the value "Student" and you can't always tell what kind of student they are; so, direct care can be helpful for that because they won't just say "Intern," or "Resident", they'll also say "Orthopedic Resident" or "Family Practice Resident" or something like that.

The Purchase Care Specialty Code is the same one that's in the Medicare claims data, and then we have the HIPAA taxonomy code like most of the people in the US would have. If you want to bill electronically, you have to have one of those.

These are just some of the code sets; I don't really--I think I said the important things I wanted to say about them, but we have dictionaries available to you so you can review those code sets before you decide what you want to use.

So, DoD EHR data. So, we talked about location, we talked about provider. Now, what we're going to do is really get more into the types of data that are collected to help you identify the kinds of care. So, first CHCS. CHCS, I started working around the Military Health System back in 1990 and we were on CHCS--we're still on it, we've been trying to get rid of it and we've made a lot of progress, but we're not quite there yet. So, much like the Vista system, the CHCS is a set of separate servers--systems on separate servers that are mostly configured alike; we try to keep them configured alike. Sometimes, timing gets in the way of that, but the idea is that they're configured alike.

So, that makes the databases smaller, but it really does cause a problem when you're trying to do clinical care because you can only see the care that is provided and documented in that local area when you're there. So, if you're in Hawaii, everybody in Hawaii uses the same CHCS, and that's great. But remember in the military, people move every two or three years, and so if somebody moves in from Hawaii, in CHCS, it looks like they're a brand-new person, we don't have any real history associated with that person, and that causes a lot of problems.

The other thing it does is it causes data quality problems. The idea that you can have these separate systems purportedly keeping the same data, it doesn't always work that way. I may have this in the slides later, but I’m just going to use myself as an example. There's a field called the "Family Member Prefix," and it tells you whether someone is the first kid, or the second kid, or the third kid, or the fourth kid. I have two sisters, so a family of three kids; and depending on what CHCS I look at, I was Kid 1, or sometimes I was Kid 2; or sometimes, I was Kid 3. So, those inconsistencies can occur for any variable that's assigned locally that doesn't get vetted through DEERS--and that was a variable that didn't get vetted. So, those data quality problems exist from host to host.

And one of the best things you can do is sort of understand that the DEERS data is going to have the best data about a person; and so if anything is DEERS or person-related, trying to get that from the DEERS data is the best.

So, CHCS, when I started, as I said, in 1990, it was the only system. But in the mid-90s, the AHLTA Project came about; and the idea with AHLTA was it was going to be an EHR--it's not quite an EHR, there are some data that are missing from AHLTA that are super important; and if you combine AHLTA and CHCS, you get all of the data, but if you just look at AHLTA, you miss a lot of data.

And I’ll go into what's--I guess there's one other EHR, it's called Essentris, and that EHR called Essentris is used for inpatient care; so AHLTA is used for outpatient care and Essentris is used for inpatient care.

There was an automatic syncing built with CHCS and AHLTA, so that if care is provided in AHLTA, that information also gets written back to CHCS; but if care is provided in Essentris, that care does not get automatically written back. So, we've been doing double data entry for inpatient care for a very long time; the data from Essentris is not available in DaVINCI; there isn't any central data store with the Essentris data on it. And what that does is there's a pretty big limitation in the data caused by this, and that limitation is--we have very good institutional inpatient data, s so the ICD-10 PCS codes and the ICD-10 CM diagnosis codes, those are very fine for inpatient care; but the CPT codes for inpatient care are greatly censored. There's a lot of stuff that just doesn't get captured.

So, I was actually--if you're familiar with Leapfrog, I was doing our leapfrog work just the other day for the MHS, and I was looking at low-volume high-risk procedures. And when you try to do case-finding for those procedures using ICD-10 codes, you get one number; and then if I go and I try to find the same impatient events, and I look for the CPT codes, they're often missing.

So, when you want to do case signing, like if you're looking for people who had hip replacements or something like that, if they were done on an inpatient basis, trying to do the case finding based on CPTs, you're going to be missing a lot of care. So, I would strongly encourage you to also bring the ICD-10 PCS code into your analysis to make sure you don't miss cases.

Here's the part where I was talking about people's data being different. I use myself as an example, so I think I won't waste time on this slide. Sorry, I went over it too quickly. This is the complex data flow that occurs to get the data to you--and I certainly don't want you to study this chart because it's super complex, but just some things to know: like everything starts in CHCS with an appointment on the top left; once you make an appointment in CHCS, that data gets sent to our MHS data repository which is your source of DaVINCI, it also gets sent through to AHLTA. And then it's just a very complex process, there's lots and lots of writing back from one system to another, and there are errors sometimes in the data, so there can sometimes be missing data because of this sort of messy process. MHS GENESIS makes this a lot better.

But two of the major failures that we would talk about, there's a thing called a write-back error--or this is one major error. It's the write-back error, it's when you capture something in AHLTA and you try to write it back to CHCS and it doesn't work. This is such an important thing that MTF commanders have to sign statements every month to make sure that they're trying to keep up with these write-back errors, they can cause some large problems especially around the time that you introduce new code sets and things like that.

At least, there is a write-back methodology with AHLTA; though, as I said, there's not a write-back methodology for Essentris. So, the doctors will be in Essentris and they'll be delivering their care, they'll be capturing everything they need to be capturing; but unless somebody comes behind them and says, "Okay, now, I’ve got to go into the administrative data, open up an appointment, record what you did," and then the data will get in the system. That, like I said, is not uniformly happening. So, be cautious with the CPT codes. But again, the ICD-10 PCS codes are fine.

I just use this picture to kind of show people the way the data flows and there's something important here to this slide as well: you can see there's communication between DEERS and CHCS; so DEERS is our enrollment data, this is where we do eligibility checks, demographic data gets updated this way, this is how PCM assignments get communicated--primary care manager. CHCS sends all kinds of data over to AHLTA; and then what AHLTA does is it sends coded professional records back to CHCS, that's that write-back. There's also information going into Essentris, but notice that's a one-way arrow, so that's what I was talking about where there has to be a double-data entry.

Anyway, since the CHCS hosts aren't connected, it's super important that we can export data from CHCS and put it somewhere essentially, so that we can do our population health work if we want to. And it's really necessary for us so that we can understand the complete healthcare history of patients and so that's why those extracts that I showed you earlier, why we really put so much focus on making sure those are available to us and to the MTFs. So, we have point, click, drag, drop systems, BI tools that the MTS can use to do ad hoc querying with all of the data that I’m talking to you about, so they can help them manage their patients more fully.

So, now, DoD source data to DaVINCI. So, this is the data flow that gets us to DaVINCI. So, what you can see here is there's this CHCS AHLTA; we have a Pharmacy Data Transaction Service, we have DEERS, we have claims, we have GENESIS, we have many, many more sources coming into what we call the MDR, that stands for the MHS Data Warehouse. That's the primary data warehouse, it's a SaaS data warehouse, there's a whole lot of cleaning done to the data, a lot of enhancements, and appended fields and things like that.

From the MDR, what we do is we create extracts and these extracts go to DaVINCI--not all of them do, but many of them do. So, what we do is we send those extracts over to a system that really it just serves as a sending system, it's not providing data for DaVINCI, but it's storing the data that's sent over from the warehouse and then from that system, HSDW, Health Services Data Warehouse, the data are forwarded into DaVINCI and then DaVINCI is a SQL database available to you guys on the VA.

So, your source, even if it sort of sounds like the data coming from this HSDW in the middle, it's actually really coming from this data repository.

And this is our Military Health System data repository is used for all the official stuff that DoD does, our reporting to OMB in Congress, and writing Congressional reports and things like that is always done from this data warehouse here. So, it's our official source; so, you're getting the data for DaVINCI from the right place.

So, in the MDR, because we have some issues that our direct care system, when they forward the data, we know we have issues, we want to do a lot of cleaning of the data so that you guys and the users of the data within the MHS don't have to do that cleaning every time they do a study. So, one of the more important cleanings that we do is a person identification enhancement. So, not all of our files have the same person identifiers; and sometimes, those things can be--the same person might be represented by more than one identifier. I know you guys have the same problem, I work pretty closely with you all trying to implement your version of Cerner Millennium, and I know this has been something you guys have been working on too and just trying to make sure that all the known person IDs are tracked for somebody, and they're de-duplicated if they need to be.

And for us, that's very important. We have people with multiple reasons for eligibility; for example, you might have somebody who's a guard reserve member married to an active-duty person, and so they're both sponsors and they have different benefits associated with their sponsorship. So, we also apply the DEERS data so that we can correct for some of the person identification problems that we get in the incoming data; we apply some coding edits because we do have some coding issues; we regroup procedures--so, we group things into diagnosis-related groups--and I should mention here our diagnosis related groups are not the same as the VA's diagnosis-related groups. So, when you look at DRGs, don't think the numbers mean the same thing across the two systems: we have a TRICARE unique grouper; we have major diagnostic categories, we have AHRQ clinical classification software, we have a whole bunch of different groupings. And there's really many, many more cleanings that we apply to this data before we make it available to users.

So, there are different DaVINCI databases. When people in DoD want to use the DaVINCI database, they go through the governance on the left side; and if people in the VA want to use the DaVINCI data, then they go through the governments on the right side, so data goes back and forth both ways--I mean it'd probably be much more efficient if it were in one environment, but that's not the way it is. But the tables are exactly the same in the two environments; we do keep them in sync.

So, here's where you can find the DaVINCI data; so, on the left side, we're showing you where all the OMOP tables are and you can see those OMOP tables sit right alongside the CDW and the CMS OMOP tables. So, think about that as--that's a huge advantage because if you wanted to incorporate data from all three systems--and by the way, there is significant overlap, I’m sure that, in all three systems. Lots of our people are Medicare-eligible, lots of fewer people are Medicare-eligible, and there's these whole lot of dual consumers across both DoD and VA. So, if you wanted to have a full picture of all of the care people got, you don't have to learn all the rules of each system if you're using OMOP, so that could be very helpful.

On the right side, though, these are all the source tables. So, these are the tables, the way they look in the MHS data repository. So, they have not been mapped to OMOP, this is the way the data looks when the DoD people tend to look at data for our own internal studies and things like that. And so, anyway, this is the directory structure so you can see where you can find this data if you want to use it.

So, here, again, is this chart that we showed you. So, if you can just follow the green lines and maybe blow this up a little bit to make it easier for you to see, but those green lines will map and show you how all of the various tables map into all the various OMOP tables. So, \_\_\_\_\_ [00:45:40], we have them available to you for FY '89 and forward, these are just inpatient stays, so records for every patient is positioned: you've got dates of care, providers, DRGs, diagnoses, and procedures, so that's an important file. All of the stays in the MHS directory system are acute care; and then we have a younger population in Medicare and that's why we don't use the Medicare DRG, so you're going to see a big difference in things like neonatal DRGs, and you'll also see a lot of our DRGs get split up between pediatric and adult; otherwise, they're pretty much the same DRG grouping system.

You can see--I think I said OB was our number one product line, and if you just look at our DRGs you can see those first five DRGs are completely related to obstetrics care, so those are the top DRGs that we perform. One thing about the sitter is that it does include observation stays, so if you wanted to remove observation stays, there's an observation stay flag that will let you do that; they're not the same as a regular inpatient.

We also have a professional services file; every record represents an encounter or an administrative record; this does include ambulatory professional service, office visits, ER, same-day surgery; it has some labs and rads that are done in clinics, and it has some physician-administered drugs. It also has inpatient professional services, but remember I told you they're not always captured very well; and then administrative records. And it's important for you to recognize the administrative records because they don't actually represent healthcare delivery.

So, here's an example of that and on the left, I’m telling you how to find them in the data. But what happens is when an MTF refers a patient to the private sector, they open up an encounter record and then they keep that encounter record open until the provider provides a report back to DHA to say, "This is what happened in the encounter that you referred." Once that happens, that happens that encounter is closed; and you can see that these are very, very large counts, so in the millions and millions of rows of this file.

And if you're just trying to count encounters, you really do want to exclude these things because it's really--the only reason they're there is that physicians get reminders to close their appointments, and so if you have an open appointment, an open encounter, they'll keep getting a reminder and that'll make them want to follow up with that private sector care provider to get the report back. So, that's why they're there, but you do want to get rid of them if you can when you're doing your studies because it's really not real healthcare.

There are up to three providers that can link to 13 procedures; we don't have the linkage in DaVINCI yet; it is in the source data and we hope to get that to you. Encounter Day, the treatment DMIS ID, the MCS code like the stop code, provider identifier, specialty, and then we have the CPT HIT PCS and the ICD-10 codes on this data.

Lab and rad files, they're available from '05. The files from '05 forward don't have results in them, but I’ll show you a different file with results. These are just CPT codes and there's lab and rad, and you have to use modifiers to tell the difference between professional and technical data.

We have pharmacy files--the pharmacy file, the source file actually has direct care and purchased care--and actually it has VA in it, so that's kind of interesting and a duplication for you guys. Whenever you fill a script for a dual eligible beneficiary, here's actually a check that is sent into this PDTS database to make sure that it's safe for the beneficiary to get the prescription; and so those VA records get sent to DoD in this PDTS file. It's not the entirety of care that the VA provides, it's just the care that the VA provides to dual eligibles.

We also have the purchased care, we have a mail order program and a retail program; and in here we have the date dispensed, we have ordering information, the appointment, the provider the national drug code, the generic class, therapeutic class, and then the source of the data as well are available. And one thing you can see, if you just look at the percentages, the VA is a pretty small percent, it's like 6 percent of the prescriptions; and if you look at the retail, that's the largest area of providing prescriptions for DoD.

We have some clinical tables; we mentioned the vital signs table. Vitals are only available if they're collected in AHLTA, so this is only office visit vitals, you're not going to see any vitals for inpatients; immunizations, as we noted, and we've got the CBX codes for the immunizations; lab results, they're stored by link and they're available in a chemistry, microbiology, and pathology and they're only available 2009 and later for all of these. And radiology results as well.

Everything is linkable. So, I know this is--it's hard to kind of read how these links go, but the idea is there's a sitter number, and a CAPER number, those are the record numbers; and anything that's ordered or associated with the event being recorded in sitter or CAPER is linkable. So, I can go into the CAPER and say, "Gee, this doctor was presented with a patient with these diagnoses and these complaints; and as a result, these are the orders they did, these are the labs, and these are the results and all of that's linkable within direct care." So, that's pretty nice. The data is going to also be linkable through MHS GENESIS, so that's not something we're going to lose.

We also have theater data. Theater data, if you want to use it, I have to tell you it's very dirty--lots and lots of hand entry, lots of fat finger problems, and also the procedure codes are almost entirely blank. So, you do get diagnosis codes in there, that’s helpful, but if you wanted to look at treatments in theater, that data is not very good for it--it's also not linkable to anything else.

Just finally, just the last part of this, a couple of the OMOP tables that I’d like to highlight. There's a Visit Occurrence table, it's got inpatient and outpatient direct and purchased care. So, the files that I just talked about, the sitter and the CAPER are in there; there's also theater. We'll talk about purchased care another day, but you can see things like the visit date, when it happened and mission source, discharge source, you can link this file to the procedures, and the diagnoses, and things like that.

We have a condition occurrence file; this is where you're going to find the diagnoses. So, what we're pulling in there are from all of the data sources that have diagnosis codes; these are going to be ICD-10 diagnosis codes coming in, but then they're getting mapped to SNOMED as the standard terminology, and this is very, very useful for identifying cohorts because you have all the diagnoses in one place rather than having to go survey all the files for a diagnosis.

Top conditions. Sometimes, the conditions don't map; one good example of where they wouldn't map would be those referral tracking encounters that I talked to you about. But you can see top conditions like essential hypertension, low back pain, hyperlipidemia, diabetes, sleep apnea, things like that. So, that's the condition occurrence table.

There's also a procedure occurrence table; and oddly enough, the procedure occurrence table includes procedures, but it also includes diagnoses that represent procedures. That's a little bit unusual and you want to make sure not to double count something because some diagnoses will be in condition occurrence as well as procedure occurrence. So, be careful with that. And again, all the data files that are CAPER that we talked, about also the ancillary files will be in there; the TMDS is going to be in there, but the CPT codes, as I said, are almost entirely blank, so there's not a lot of value in what's provided there.

If you look at the top procedures, this unlisted evaluation and management code is something that... it doesn't mean anything, the Code 99499 in DHA is just a placeholder. And then there's office visits, and health examinations, and things like that. That health examination of a subgroup is actually where we're putting encounters for post-deployment health assessments and reassessments and things like that.

There's a drug exposure table. So, this is not just coming from the pharmacy data; we have a CDR medications table that looks at inpatient meds; there's immunizations that feed into this and then, sometimes, drugs are represented by things that you get for an infusion or something like that that would be identified in the CAPER, so that's reading into it the ancillary data and even the TMDS data; and here's where you can tell what kind of meds people got, how they got them, when they got them, who ordered them, and stuff like that.

Top drugs. Ibuprofen, anthrax vaccine, et cetera. There's the influenza vaccine, I see on there--hopefully, we'll see a COVID vaccine showing up real soon. There's also the measurement table, this is primarily where you're going to get clinical results, so chemistry results are a big deal; vital signs are showing up in this file, blood pressure shows up in this file, stuff like that.

So, we mapped things according to the OMOP data; and then in cases where they didn't have OMOP concepts to exist, we created our own mappings for some of the stuff in the vital sign data. Everything's really well populated in the data dictionary so you can get descriptions from that. But see how this is kind of a mixed bag: I can see lab data in there like CBCX, and I can see links and CPT codes; and there's blood pressure, height, weight, heart rate, et cetera. So, a lot of good clinical data. There's a note table, this is where we're going to put the radiology reports, the microbiology reports, and the pathology reports, so you can do free text mining on that--I’ll tell you those kinds of reports, right now, are not reports though that they're giving us from MHS GENESIS--and I hope you guys are luckier than we are to be able to get that data when you guys get to Cerner Millennium.

Finally, there's just this mixed bag table, this observation table and we just kind of put stuff we thought was interesting in here so things like tobacco user or major diagnostic categories are in there, or whether someone got alcohol screening--just a lot of different mixed-bag stuff in that observation file.

So, just summarizing, we've covered the sources, the processing flow, we talked a little bit about completion, we gave you an overview of the direct care and talked about how you can link everything. I talked about the limitations of inpatient professional coding and we also really wanted you to take home the point about the non-clinical encounters being in the data, so you would know to remove them if you needed to remove them.

I provided you some useful links here, so if you look under the data sources, the section on VINCI Central, you'll be able to find our data dictionary, our mapping design documents, and all the ETL logic. And then there's some data training and stuff like that at the DaVINCI data academy.

And then, that's what I wanted to cover today; I know I really took quite a lot of time, but I’m happy to answer any questions if you have any. So, thank you very much for your time and attention.

Rob: Thanks, Wendy. Yeah, we're just about out of time; but we do have three questions queued up, so we should be able to get through them.

Wendy Funk: Perfect.

Rob: First up: "How were the market area boundaries drawn? Are they based on DoD data or metro area data drawn from the US census?"

Wendy Funk: So, they're completely based on DoD data. What we're really doing is we looked at all of our large hospitals and decided to draw markets; if they had smaller hospitals around them then we would draw markets around them so they don't correspond very well to anything that you're going to find in the private sector data; they're going to be very specific to DoD.

Rob: Thank you. "Is the type of visit, specifically telemedicine, captured?"

Wendy Funk: So, there's not a flagged variable for telemedicine, but that's a very important topic right now, and our telemedicine data is recorded using CPT codes and CPT code modifiers. And so, that is something probably we should discuss as a team trying to find a way to get you guys a flag variable. But to whomever sent that question, I believe I put my email address on my first slide and I would be more than happy to provide you with the specific details about what codes to use in order to capture telehealth.

Rob: Thanks. Maybe you can go to that first slide so that the person... I’m sorry, what?

Wendy Funk: Somebody said go to the first slide?

Rob: I was just asking you maybe you could go to that slide where your email address is now so they could write it down.

Wendy Funk: No, it's not there. I’m sorry it's not there. I can put it in the chat, can't I?

Rob: You can send it to me and I’ll read it off. I think, yes, you can send it to the chat to all attendees or everyone. While you're doing that, "For a new VA clinical researcher wishing to access a list of patients with a rare disease in order to subsequently provide care for a pre-to-post quality improvement project across a VISN, what is the best process to navigate governance?"

Wendy Funk: Oh, that's a great question. So, we have an MOU between DoD and VA that allows each agency to use their own governance. So, if you want to use DaVINCI data for that purpose--and I think it sounds like a great purpose, especially if you're looking for something rare, those diagnosis and the condition occurrence and procedure tables would be really good for that, you just have to go through your VA governance. So, I don't typically go through that governance process, but there shouldn't be anything different about it. They're able to approve your use of DoD data; you don't have to get a separate DUA with DHA any longer if you use this process.

Rob: Thank you. While we were questioning and answering, or one more came in. This person asks, "Can you share the link to the VINCI training academy?" I can tell people what your email address is if the person wants to email you.

Wendy Funk: Yes, that'd be fine. I don't have that link off the top of my head. I’m sorry about that.

Rob: So, Wendy's email address is wfunk@kennelinc.com; that's K-E-N-N-E-L-L-I-N-C.COM. wfunk@kennelinc.com.

And with that, Wendy, do you have any closing comments before I end the webinar?

Wendy Funk: I just want to thank you for your attention; I think that this is a great opportunity to have these two departments together as one. I, as a DoD researcher from the DoD side, have often struggled to get access to VA data and I’ve heard from you guys the same thing over the years that I’ve been working here. So, we really would like you to use this data. And thank you so much for your time.

And again, if you have questions, just shoot them to me. I’m happy to help.