Cyberseminar Transcript

Date: January 9, 2020

Series: VA Informatics and Computing Infrastructure

Session: Medicare Data in the OMOP Common Data Model

Presenter: Kristin de Groot, MPH

*This is an unedited transcript of this session. As such, it may contain omissions or errors due to sound quality or misinterpretation. For clarification or verification of any points in the transcript, please refer to the audio version posted at* [http://www.hsrd.research.va.gov/cyberseminars/catalog-archive.cfm](file:///C%3A%5CUsers%5CVHASLCMyersK%5CAppData%5CLocal%5CMicrosoft%5CWindows%5CTemporary%20Internet%20Files%5CContent.IE5%5C73RXHSNF%5Cl)

Rob: And as we are just now at the top of the hour we’ll go ahead and get things started. I’d like to turn things over to Kristin de Groot who’s with VIReC on a joint project with VINCI. Kristin can I turn things over to you?

Kristin de Groot: Yes. Okay. Can everyone see this?

Rob: We can, thank you.

Kristin de Groot: Okay. Thank you. So as Rob mentioned I’m Kristin de Groot from VIReC, the VA Information Resource Center. And today I’ll be speaking about our efforts to transform Medicare data into the OMOP Common Data Model.

So before we get started I’d like to recognize our team. So as Rob mentioned this is a partnership between VIReC in our role as the data steward for Medicare data for VA research use. And VINCI who’s leading the transformation of VA data into the OMOP Model.

So here’s a brief outline for my talk today. First I’ll give brief overviews of both Medicare and the OMOP Common Data Model and then I’ll talk about what the Medicare data look like in the OMOP Model and give an update on the current status and highlight a few of the challenges we’ve encountered. In the gray box here is for the OMOP, the VA OMOP data and I won’t be discussing this data today but I wanted to include it on this slide because the last topic I present on is using the Medicare OMOP data and VA OMOP data together.

So now we’ll pause for two poll questions to the audience. The first is about your experience with Medicare data.

Rob: Kristin that poll is up, excuse me. The question being please rate your experience with Medicare data. The software would not let me leave any of these blank Kristin so I took it upon myself to write in a little knowledge, some knowledge, and a lot of knowledge for the middle three choices.

Kristin de Groot: Okay.

Rob: I didn’t think you’d mind. We have about 70% of your viewing audience having made their choices and it usually levels off right around 70-80% so I’ll give people a few more moments to make their decisions. And yeah it does seem to have leveled off so I’m going to go ahead and close the poll and share out the results. And let you know that 23% of your viewing audience say they have no knowledge, 44% the largest number say that they have a little knowledge, 19% some knowledge, 14% a lot of knowledge, and nobody says they have expert knowledge.

Kristin de Groot: Okay.

Rob: Would you like to launch right into the second poll.

Kristin de Groot: Yes please.

Rob: Okay. And that’s up. Kristin would like you to please rate your experience with OMOP Common Data Model as opposed to Medicaid. Again the same choices no knowledge, a little knowledge, some knowledge, a lot of knowledge, and expert knowledge. And again we’re up around 67, almost 70% so we’ll give people just a few more moments to make their choices. Great and it looks like people got them in at the last second so I’m going to go ahead and close the poll, share out the results. And Kristin 41% say they have no knowledge of the OMOP Common Data Model, 29% say they have a little, 16% say some knowledge, 14% a lot of knowledge, and again nobody says they have expert knowledge. Except you of course! Now we’re back on your slides.

Kristin de Groot: Okay. Thank you! So that poll information was very helpful. So it really reinforces the need for these two introductory sections; intro to Medicare and intro to OMOP. So first an intro to Medicare. I think it’s important to have a basic understanding of our source data and why it’s important to VA research before we talk about what it looks like in the OMOP Model.

So among Veterans under 65, 12% are enrolled in Medicare. And people under 65 can be eligible for Medicare due to either disability or end-stage renal disease. Among Veterans 65 or older almost all are enrolled in Medicare. If a person is eligible for Medicare due to disability, end-stage renal disease, or old age they can enroll in Medicare regardless of their income or whether or not they have access to other health insurance like Access to Care at the VA. And throughout this presentation when I use the term Veterans, I’m referring to Veterans enrolled in VHA health care.

So next I want to highlight the fact that there’s a variety of types of Medicare coverage. The first is Fee-for-Service which is sometimes called original or traditional Medicare. And in this option coverage is administered directly through the Centers for Medicare and Medicaid Services or CMS. The second option is to enroll in a Medicare Advantage Plan or Managed Care Plan and in this option beneficiaries can choose from a variety of plans and enroll in a plan that’s run by an insurance company that’s contracted with CMS to provide benefits. And among Veterans about 25% are in Medicare Advantage as opposed to Fee-for-Service.

And among those who enroll in Fee-for-Service they can choose Parts A, Part B and/or Part D. Within Medicare Advantage the plans are available with or without Part D. And from here on out in the presentation I will primarily be focusing on the utilization data generated from people with Fee-for-Service Parts A and B.

So I want to talk briefly about the sources of VA and Medicare data at a high level and how the different sources result in different data. In the VA a Veteran obtains care at a VA facility and the data is entered into VistA and this data makes its way into databases like the VA’s Corporate Data Warehouse, CDW that can be used by researchers. But if a patient sees a provider outside the VA the provider will still likely enter data into an EHR but this is not what we consider Medicare data. Because Medicare is a payer not a provider the data don’t get to Medicare until a claim is submitted. Medicare data are claims or bills. So providers submit claims to Medicare and this data is then made available to researchers. But because the Medicare data are based on claims there is little to no clinical data, like vital signs or lab results since these aren’t needed for billing. And I also want to mention that Medicare data are not included in VA databases. If you want to know what services were received outside the VA and paid for by Medicare, you need to use the Medicare data.

So in Medicare there are two types of providers. And each of these providers uses a different type of claim form or bill. And because each of the claim forms collect information differently we would end up with two different types of data. So one type is institutional providers such as hospitals, nursing facilities, home health agencies, or hospices. And the data from these bills ends up in one of five institutional files. And these are inpatient, skilled nursing, home health, hospice, and outpatient. The outpatient file here contains claims submitted by institutional providers like hospitals but the care was in an outpatient setting. The other type of providers are non-institutional providers. And these include a wide variety of providers including individual physicians. And this data ends up in one of two non-institutional claims files. The carrier file contains claims from most of these providers including physicians. And durable medical equipment is mostly claims from medical suppliers.

So most of the time a claim will contain one visit. Such as one acute inpatient stay or one office visit. However sometimes a claim will contain multiple visits. And most often this happens when there are multiple visits to the same provider, for the same purpose, in a short period of time like physical therapy, dialysis, or home health care. And in this situation claims will need to be split apart to accurately count the number of visits. So on the other hand sometimes a single visit will be spread across more than one claim. An example of this is emergency room visits where both a hospital and a physician may bill separately for each of their services. And another situation where we see a single visit or stay being represented by more than one claim is with long stays, especially those in skilled nursing facilities.

So here’s an example of a stay, a Medicare stay that’s made up of two claims. And then this could be in the inpatient stays or skilled nursing facility stays. And the first half of the claim covers the first half of the stay and the second claim covers the second half of the stay. And in some cases this might have been submitted as one claim but the facility can choose to split the stay into multiple claims so that they can get reimbursed more quickly or it makes for easier accounting for them to end the claim at the end of the month or the year. So CMS realizes it’s extra work to transform this data from claim level to stay level so they created the MedPAR file.

And this file is created from the inpatient and skilled nursing claims but claims are rolled up to the stay level, that’s the technical word. In addition to being easier to use another benefit of the MedPAR file is that it contains some stays from the Medicare Advantage enrollees. Those stays are not included in the claims’ files. However when CMS converts the data from claims to stays not all the information from the claims make it into the MedPAR file. So I want to highlight this MedPAR versus claims issue because it’s one of the, it’s an example of a decision that researchers need to make when using the Medicare data and it’s one of the decisions that we had to make when converting Medicare data into the OMOP Model which I will talk about later in this presentation.

So next we’ll talk about the OMOP Common Data Model at a high level.

But first I want to take a step back and just talk about a data model. What is a data model? It organizes data elements and defines how they relate to one another. And I know you can’t see the details here but this is an example from the VA CDW and this is the data model for the Inpatient 3.0 Domain. And each of the boxes here represents a table and the lines show how the tables relate to each other.

So if you’ve ever done research that uses data from more than one source you know that combining different data types can be challenging. And an example shown here you have the VA data like CDW which is based on data from the Electronic Medical Record and as we learned Medicare data is based on billing data. So you might also be using other types of data, maybe other data from outside VA and all these sources have their own way of organizing the information. So transforming data into a Common Data Model can help with this problem. By transforming all the sources into a standard format they can more easily be used together.

So as we saw in the last slide one of the biggest benefits to a Common Data Model that it standardizes how data looks. Most visibly through the names and contents of the tables and the relationships to other tables. But also through standardization of variable or field names and the values contained in those fields. Another benefit of Common Data Models is the ability to incorporate or embed knowledge of underlying data into the model. Almost every dataset will have nuances that users need to be aware of and these can be taken into account when the source data are transformed into the Common Data Model.

So after deciding to use a Common Data Model the next question is which one to use. And there are several options out there but we decided to use OMOP, the Observational Medical Outcomes Partnership. So why OMOP? Well for us, honestly the main reason we decided to use OMOP with the Medicare data is that it’s already in use in the VA with the CDW data and it recently became available for the Department of Defense’s Military Health System data. And beyond that it’s used by more than a hundred health data systems around the world. So why are all these groups choosing OMOP? Compared to other data models it has broad coverage meaning it captures a lot of the information from the source data. Another benefit is that the OMOP user community has developed a lot of open source tools including code and software that can be used on the data that’s been converted to this format. And lastly OMOP relies heavily on the use of standardized vocabulary and concepts.

So a little bit more about that. One of the key features of OMOP is the concepts. The OMOP Model has a whole table of concepts that contain almost all code sets that you might need to represent health care data. Like some, just a few of the examples ICD-9 diagnosis codes, CPT codes, NDC codes, and many, many codes you’ve probably never heard of. There are over six million of these codes, these concept codes. Each one of these is represented by a Concept ID which is a unique number that isn’t used within OMOP to represent anything else. So you’ll never need to ask is this an ICD code or is this a CPT code.

So here’s an example with code 25000. And unfortunately in the Medicare data all the ICD codes have been stripped of their decimals. So if you aren’t careful you could possibly confuse a code for diabetes with a code for a wrist incision. But in OMOP each of these codes will be assigned a unique Concept ID so there’s no confusion about what code is being used.

A second feature of the standard vocabularies I want to highlight is the idea of standardizing concepts. And here’s an example using drug data. The types of source codes are listed in the lower half of the diagram and some commonly used vocabularies when dealing with drug data in the VA are NDC codes or VA product codes or especially if you’re using the Medicare data the HCPCS codes, HCPCS. And the OMOP Model converts all of these codes to a standard vocabulary. And in this case the standard is RxNorm. So no matter what vocabulary your source data uses you have the ability to search for drugs using just one code set.

So there are two ways to search OMOP’s vocabularies and how these, all the concepts relate to each other. The first is a web-based tool called Athena and you can search using the web interface or you can download all the vocabularies from this site. And the second option for users in the VA is to use the SQL tables that you receive when you obtain the rest of your OMOP data on VINCI, either the VA, the DoD, or the OMOP, I’m sorry the Medicare OMOP data. And if you use the SQL tables that come with your OMOP data on VINCI there will be some codes that aren’t found in Athena and that’s because the groups in the VA who transform the data into OMOP have added some custom codes that aren’t currently found in the OMOP vocabularies.

So here’s a diagram of the OMOP Common Data Model. And each of the white boxes represents a table and the arrows describe how the tables relate to each other. And here you can see there are ten standardized vocabulary tables and the one on the top there is the concept table which has the six million concepts.

And here’s another representation of those standardized vocabulary tables and you may not be able to see all the table names here but the important thing to note is that the concept table are central to all the other nine standard vocabulary tables. And the concepts from this table are used in pretty much all of the other tables in the OMOP Common Data Model.

So that’s the high-level view of OMOP and now I’m going to talk about Medicare data in the OMOP Common Data Model.

So as you remember in this diagram each of the white boxes represents a table. And in the transformation of the Medicare data we aren’t currently using all of these tables. So I’m going to walk you through the tables that are being populated by the Medicare data.

So the first three tables I’m going to highlight fall under the heading of standardized health system data, in the pink box. Location, care site, and provider. If you’ve used CDW data these would be similar to dimension or dim tables. They don’t contain any patient-level data. The first is location which contains a record for each unique location. For facility locations we have a full street address that we obtained from the publicly available Provider of Services file. And for patient locations we have data only as detailed as the zip code. We don’t have the full address. Which we obtain this from the Medicare Vital Status file.

Next is care site which includes a list of places where care can take place. And we created this file by combining all care sites found in the MedPAR and the claims files. Examples are hospitals, skilled nursing facilities, or dialysis centers.

And the third standardized health system table is provider. Which is one record per individual health care provider. So this is physicians and other individual people who bill Medicare. Similar to the care site table, this table was created by combining all individual providers found in all of the Medicare claims. And we added some data from the publicly available NPI file for additional information about the providers.

So now we’ll shift to the blue box on the left, starting at the top with the person table. And this table is really needed to properly use any of the other tables. Almost all of the other tables link to the person table. It contains demographics and a link to the person’s location, in the location table. And most of the information in this table comes from Medicare Vital Status.

So next, observation period table. This tells you the time period for which data could be found in the model for a person. So this is the table that probably has the most variation in terms of how it’s populated depending on what kind of source data you’re dealing with. In the case of Medicare data in the VA we choose to use this table to express the years for which Medicare data could be available in the VA. And in some cases this might be different from the years that a person was enrolled in Medicare. This is because the VA submits a cohort of Veterans SSNs to CMS each year, which we call the VHA cohort file and will only have Medicare data for a person if their SSN was submitted to CMS for that year. For the most part we have CMS data for Veterans from the time they first enroll or start using VA health care or start receiving VA compensation or pension benefits. An example, if a person was enrolled in Medicare starting in 2000 but they weren’t enrolled in the VHA until 2005 we would only have Medicare data for that person starting with 2005. So in this table this person’s observation period would begin in 2005.

So the next table I want to talk about, payer plan period which is technically one of the health economics tables but I think it’s important to talk about along with person and observation period. This will tell you if or when a person was enrolled in Medicare, whether they were in Parts A, Part B, or Part D and whether they were enrolled in a Medicare Advantage plan. This information is sourced from the Medicare enrollment files called the Denominator or the Medicare Beneficiary Summary File, MBSF. And this is important so you know whether or not there might be utilization for an individual, utilization data.

So going back to the box on the left we have visit occurrence. This is where you find visits, stays, and encounters. Examples of types of visits include inpatient stays, long-term care stays, outpatient visits, emergency room visits, or emergency room visits combined with an inpatient stay. And the visits in this table are sourced from the MedPAR and the claims files.

The visit detail table is new to OMOP Version 6 so not all instances of the OMOP Model include it yet. I don’t think the VA OMOP data yet includes this table. It includes more details about what happened during the corresponding visit occurrence.

And lastly we have what I call the code tables. That’s not the official name but that’s what I like to call them. Condition occurrence, drug exposure, procedure occurrence, device exposure, measurement, and observation. The information in these tables comes from the diagnosis and procedure codes in the claims data. Earlier I mentioned that in the OMOP Model codes are transformed to use standard vocabularies. And these tables include both the original code called the source code and the standardized code. And while these tables contain information from diagnosis and procedure codes this is obviously not how the information is organized. We don’t have all diagnosis codes in one table and all procedure codes in another table. When the source codes are transformed to standard codes the model indicates what type of code it is and what table it belongs in. Some diagnosis codes aren’t really diagnoses but they’re observations. And similarly some procedure codes aren’t really procedures but they’re drugs or devices.

So here you can see the relationship between the source codes found in the Medicare data and their final destination in the OMOP tables. And while all ICD procedure codes end up in the procedure table. ICD diagnosis codes, CPT, or HCPCS codes can end up in almost any of the code files.

So this table shows the most frequent code found in each of the OMOP code tables based on the data we’ve transformed so far. And some of the tables are pretty self-explanatory but I do want to talk a bit about measurement and observation. Earlier in the presentation I mentioned that Medicare data don’t include clinical data like lab results. But in the measurement table we do find fact of lab tests, right here. Also in this table there’s a field for measurement value and in this case it would, it should be the lab results. But you’ll see here that for 98% of the measurement records the value is missing. And even when it’s not missing the most detail we get usually is something like abnormal. So it’s not really useful if you’re interested in lab results. And the second table I want to draw your attention to is observation. And this table is a bit of a catchall for things that don’t fit into the other tables. Like measurement this table has a value field but it’s used a little differently than measurement. The most common value is history of clinical finding. And so in this example you need to look at the value field to find out what that finding was. Almost half of the rows in this table are missing an observation value. Not because the historic clinical finding is unknown but just because a lot of observations in this table don’t require a value. I’ve also listed another example of a value here to give you an idea of what kind of values are found in this table.

So next I’m sure this is the main reason some of you joined the Cyberseminar today, what is the current status of the data.

Version 1 of the data was recently released to a few projects to begin alpha testing. And we hope to make this data available to other approved research projects in the next few months but the timeline is dependent on what kind of feedback we get back from those alpha testers. And the Version 1 of the data is based on calendar year ’99 to 2017 Medicare data. The tables person, observation period, payer plan period, location, and provider are complete. Meaning they contain all data for this timeframe. But this version of the data does not contain all visits and visit related information. It only contains inpatient and skilled nursing facility stays.

We are currently finishing internal testing of Version 1.1 and we hope to make this data available to our alpha testers in the next month or so. Again availability to all projects will vary depending on the feedback we receive from the testing projects but we hope that it will be available in the next few months. And Version 1.1 will include everything from Version 1 plus institutional outpatient visits. And these are visits from claims submitted by institutional providers like hospitals but were in an outpatient setting.

So here are some record counts from Versions 1 and 1.1 of the data. You can see that our person and observation period tables include about 20 million people. But not all of these are enrolled in Medicare. Only around 12 million were actually enrolled in Medicare. The other 8 million are people whose SSNs were submitted to CMS but they weren’t enrolled in Medicare during our time period. Version 1 of the data includes over 35 million visits and that grows to almost 300 million after we add the institutional outpatient.

So our next steps will be to add the calendar year 2018 data which have just recently arrived in the VA and our next priorities will be to add the Medicare carrier claims which are claims submitted by physicians, labs, and other noninstitutional providers. And we will also work on adding the Part D prescription drug events. When your project does receive Medicare data in the OMOP format it’s very important to know what types of visits are or aren’t included in the data. Because as I mentioned currently not all types of Medicare visits are in the OMOP data. And it will probably be the end of 2022 before all visit types are included. As listed here carrier claims and Part D events aren’t yet included. Other types that aren’t yet included include home health, hospice, and durable medical equipment. So if you want to use Medicare data in the OMOP format in the near future you will likely need to use some data in the OMOP format and some data in its original format.

So next I want to highlight a few of the challenges or issues we encountered when transforming the Medicare data into OMOP.

So whenever anyone uses data there’s going to be decisions that need to be made along the way. Like dealing with invalid or missing values. Figuring out how to deal with conflicting information from different sources or just how to handle things in the data that don’t make sense. When we’re doing this transformation with the Medicare data we needed to think about our decisions not from the point of view of one project but how we can maximize its usability for the largest number of projects, while also maintaining data integrity meaning we don’t want to inadvertently create data or make inferences that didn’t exist in the source data. But we also don’t want to potentially exclude important data from the source files. On the other hand we’re dealing with millions and in some cases billions of records. The logic decisions we make have to be scalable. We need to try to avoid developing logic and code that would be used only on a small number of records or we’ll never finish the transformation. We need to be able to clearly communicate our logic decisions so that if needed the transformation could be replicated by users inside or outside of the VA. And all of this is a balancing act. Our team makes decisions with all these factors in mind but as we and other users start to work with the data we may see the scales are tipped too far to one side or the other. And that in the future we may decide to make changes to this transformed data. So next I’m going to go over a few specific issues we encountered.

The first of these challenges is related to the person ID, the OMOP person ID. In the Medicare data a person’s unique identifier is the real SSN which is then scrambled when it comes into the VA. In the CDW a person’s primary identifier is the ICN although other identifiers are also available. In the OMOP Model it’s recommended that a randomly assigned person ID is used. Our goal when transforming the Medicare data was to use the same person ID as is used in the VA OMOP data and the DoD OMOP data.

So the first problem we encountered is that not all Veterans in the Medicare data are found in CDW. So therefore they wouldn’t be in the VA OMOP. We found 1.7 million scrambled SSNs in the Medicare data that weren’t in CDW. And most of these Veterans are VA compensation and pension recipients. So for these Veterans we created a new person ID but we had to be careful not to overlap or to someday overlap with person IDs created by the VA OMOP team. So we started with a high number 50 million is the first person ID unique to the Medicare data. And we also realized that probably annually we’ll need to reevaluate these Medicare specific person IDs since each year new Veterans enroll in and start using VA care. And we want to make sure that when the VA OMOP team assigns a new OMOP person ID to these Veterans that we’re using the same person ID in the Medicare OMOP data and the VA OMOP data.

So a second issue we encountered related to person ID was the fact that the scrambled SSN and the CDW’s ICN were not one-to-one. And this was a problem only in a tiny percent of our 20 million subjects but it was an important problem that we had to deal with nonetheless. So we found about 70,000 scrambled SSNs associated with more than one ICN and therefor more than one OMOP person ID. So after reviewing a small number of these records we decided to choose the smaller of the ICNs, a simple solution. We also found 300 ICNs that were associated with more than one scrambled SSN. And the solution here was more complicated and we actually had to come up with several different solutions depending on which, if either of the SSNs was found in the Medicare data.

So a different type of challenge we had to encounter was whether we should use MedPAR or the inpatient and skilled nursing claims in the OMOP transformed data. So earlier in the presentation I mentioned some of the differences between MedPAR and the MedPAR stay level file and the inpatient and skilled nursing claim level files. And there are pros and cons to each of the files. And individual research projects can choose the file that best meets its own needs but we really had to think about what would meet the needs of the largest number of projects.

So our first step in deciding what file or files to use was to compare the files. While not exactly to scale, this diagram shows that the MedPAR file which is the larger blue circle contains almost all the claims which are the smaller red circles. The purple areas are the claims that have a corresponding stay in MedPAR and we saw a very small number of claims that had no corresponding MedPAR record. And those are like the pink crescent shapes on the outside of the blue. And there are some stays in MedPAR that aren’t found in either of the claims’ files. And those are the areas in blue. And these stays are mostly inpatient encounters from people in the Medicare Advantage plans which we didn’t expect to find in the claims.

So based on what we found in our MedPAR claim comparison we decided to use MedPAR as our gold standard for the inpatient and skilled nursing stays. We use this file for the fact of stay including admission and discharge dates and the facility information. We will drop the claims with no associated MedPAR record given that the MedPAR file is created from inpatient claims, inpatient and skilled nursing claims there were relatively few not in MedPAR. And because the MedPAR summary file doesn’t contain all the information found in the claims files we used the claims files to add provider or physician information and to obtain all diagnoses and procedure codes.

So here’s just a summary of our decision. This shows the stays that are and aren’t included in the OMOP Model. About a half a percent of claims weren’t found in MedPAR and so therefore won’t be included in the Medicare OMOP data.

So finally I’m going to show two simple examples using Medicare OMOP data in combination with VA OMOP data. And in both of these examples the take-home message isn’t so much the results I show but the fact that having the data in this common format makes the analysis easier to do.

And the first example is about person-level data. Race/ethnicity is unknown for about half of Veterans in VA data. So how can the Medicare OMOP data be used to easily fill in race and ethnicity data missing from the VA OMOP data?

So here we see the most frequent combinations of race and ethnicity in the VA OMOP data. And a few things to note here. The fields Race Concept ID and Ethnicity Concept ID are utilizing the OMOP Race and Ethnicity Concept codes. I’ve added the words in parenthesis just so the audience knows what each of the codes stands for. And here I’ve highlighted the rows with the missing race data, about 51%.

So because we’re interested in both race and ethnicity from both VA and Medicare person tables we’ll join these tables on person ID. Both of these tables contain many fields but we’re really only interested in Race Concept ID and Ethnicity Concept ID from each table. So the resulting table will have one record per person and will contain fields from both of the source tables.

So our primary result here is that we were able to reduce the number of Veterans with missing race and ethnicity data to about 29%. But a few more things to point out in this table that is very busy. You’ll note that for about half of Veterans there is no race or ethnicity Medicare data because they aren’t enrolled in Medicare as represented by these empty boxes. Another thing to note Medicare collects race and ethnicity data as a single variable. So people can’t select both a race and an ethnicity so that’s why we have so many people missing ethnicity data in the Medicare data. And this is a limitation of the source data so converting the data to a Common Data Model doesn’t solve a problem like this. And for some of the other rows you see that the same values are used in both the VA and the Medicare OMOP data which makes them easier to use together.

So my second example uses visit data. And we know that if you use only VA utilization data you may not have a complete picture of health care use. Veterans can be hospitalized in a VA Medical Center, in non-VA hospital paid for by VA which is sometimes referred to as VA Fee Basis, or they could be in a non-VA hospital paid for by Medicare. And of course there are other payers but we’ll stick to these for this example. So among Veterans how many hospitalizations are in a VA Medical Center through Fee Basis or through Medicare. And how can having data in the OMOP Model help us easily answer this question.

So as opposed to adjoin as we did in the last example we will union the visit occurrence tables from VA and Medicare OMOP. The resulting table will have the same number of fields but will contain all rows. And in this example we want to use several filters or where statements. We want to limit the type of visit represented by the Visit Concept ID to inpatient visits which is OMOP Concept ID 9201 and emergency room visits followed by an Inpatient Visit Concept ID 262. Currently the VA and Medicare OMOP data have slightly different timeframes so we’ll need to choose a timeframe where we have complete data in both systems. So in this example we’ll use calendar years 2000-2017. And lastly we want to limit this analysis to Veterans in CDW and we can do this by limiting the analysis to person IDs less than 50 million.

So the results of this show that about 75% of hospitalizations among Veterans are found in the Medicare data. And while this result is interesting in and of itself the point that I want to make with this example is that I was able to do, to run this analysis with only about ten lines of code and it finished running in about ten seconds. Which you couldn’t do if you were using the CDW in its original format and the Medicare in its original format.

So finally a few resources including how to request the data.

So the VA/CMS data for research project is a special project based at VIReC and is the data steward for all CMS data used for VA research. We distribute Medicare and other CMS data to VA approved projects. And we provide assistance to projects using CMS data. And all of this, including the data is no cost to researchers, to VA researchers.

So information about requesting CMS data including links to the required forms can be found on the page shown here on the left. And the link is at the bottom of the page. Even though the Medicare OMOP data is in a new format it’s still CMS data which is requested through and approved by VIReC. One of the forms that is required is the VA/CMS Data Description form which is shown on the left here. And this is the form where researchers indicate which specific CMS files and years they’re requesting. Currently the Medicare OMOP data aren’t listed on this form but once it’s available it will be added to this form. And we will make an announcement to current CMS data users and to the HSRData Listserv when that data become available. If your project, after your project is approved you will receive the Medicare data, I’m sorry the Medicare OMOP data in SQL tables in your project’s VINCI database. But the data will be in separate tables from the VA OMOP data.

Here are a few resources if you want to learn more about CMS and Medicare data. The first two links are to VIReC and specific to the Medicare data in the VA. And if you want to know more about Medicare data I will be doing another Cyberseminar next month that will go into more detail about using Medicare data in VA research. Both the VIReC website and HSR&D website will have information about registering for this soon. It’s not up there currently. And the two links on the bottom are external to VA and these are resources for all users of CMS data inside and outside VA.

Here are some resources if you want to learn more about the OMOP Common Data Model. These resources are all maintained by odyssey, that’s how you pronounce the acronym OHDSI. This is the organization that currently maintains and updates the OMOP Model and the standard vocabularies and the tools used with the OMOP data.

And lastly here are some resources for learning about the VA data in the OMOP Model, referred to as VINCI OMOP.

So that is all I have for today. And now I will take some questions.

Rob: Thanks Kristin. We do have a number of questions queued up. But let me take the opportunity to let the audience members know that if you do have a question there’s a question section in the GoToWebinar dashboard. You can actually pull that right out and expand it. You can see what you’re writing even better. And with that I’ll launch right in. The first question, are doctor’s offices included in the care site table?

Kristin de Groot: They will be. So far the transformations we’ve done have been institutional providers and so we haven’t gotten there yet but yes they will be.

Rob: Thank you. The next person asks, what is the best way to obtain lymphocyte values during a treatment period?

Kristin de Groot: So if you’re using the Medicare data you may not be able to get that information from the Medicare data. Because as I said we don’t have any lab results. So I think your best bet is just to get it from the VA data and if that result can be expressed as a diagnosis code you might be able to obtain some information that way.

Rob: Thank you. These next two questions which are the last two that we have at this time, but I would imagine people will write in more, are asking about race. The first one, this person asks what about folks with more than one race? And this was towards the end of your presentation. If that helps.

Kristin de Groot: Yes, let me, let me just go back a few slides. So in the Medicare data people cannot choose more than one race. It is, there is only one option. In the VA race data I don’t know exactly how they handled this it could very well be that there is an option for multiple races but it’s just not shown on this screen because this screen shows only the most frequent combinations found in the data.

Rob: Thank you. The next one again is about race. They say Concept ID 8552 which you have up on your screen now is used for both ethnicity and race though it is a race concept. Does this happen in other types of fields? And in parenthesis they write mixing of Concept IDs.

Kristin de Groot: I’m not going to try to answer that question because that is how it is done in the VA OMOP data. I’m not sure, I can’t answer why the decision was made to use race unknown instead of ethnicity to unknown. I know in the Medicare data because we have only, people could choose only a race or an ethnicity. If people selected a race we left the ethnicity as zero which meant there was no data. But the question about the VA ethnicity unknown that would be a question for the VA OMOP group.

Rob: Do you by any chance know what that email address is Kristin? Would it be vinci@va.gov?

Kristin de Groot: Yes. You could send it to that email address and it will get sent to the correct people.

Rob: Thank you. Will the CMS OMOP tables be available as a table or a view like the other OMOP data?

Kristin de Groot: They will be available as views. You would receive them as a view in your project’s SQL database.

Rob: Thank you. Does the Medicare data have any medication information?

Kristin de Groot: Yes. Let me go back a few slides. So some of our priorities for the next few months include the Part D prescription drug events. So this is where prescription drugs will be found for people enrolled in Medicare Part D. But it, it’s not\_

Rob: Thank you.

Kristin de Groot: \_it is not included at the data at this time but it will be coming.

Rob: This last one is a helpful comment, actually it’s not a question. This person writes, for help finding general OMOP lab results please email helpdesk concierge at vinci@va.gov. Thank you for that.

Kristin de Groot: Yes. Any, any questions you have about the VA OMOP or the VINCI OMOP data can be sent to vinci@va.gov.

Rob: Well that was the last question we have at this time. Although there still is time so it’s possible that, well another one just popped in. Pardon me. What years will medication data be available for?

Kristin de Groot: Starting with 2006. That is when Medicare Part D went into effect so we will have data from 2006 through, initially might be through 2017 but it will depend on the timing. We might go with 2018 right off the bat. But yes 2018 would be the most recent data that’s available. There is a delay in the Medicare data compared to if you’re used to working with CDW data which is close to real-time.

Rob: This next one just came in. Does Medicare data include costs?

Kristin de Groot: It does. We have not included the cost data yet in the transformation. Let me go back to the, one of these slides here. So it does include costs down here but yes that is another limitation to our initial releases, is that we have not included the cost data yet. So if you need cost data you can use the original source Medicare files which can also be obtained from VIReC.

Rob: Thank you. The only questions that we have at this time are regarding the slides and if you registered ahead of time you received an email approximately four hours ago with the link to join the webinar and it also had a link to download the slides. Additionally in two days you’ll receive a follow-up email with a link to the archive of this webinar where you will have available to you audio recording, video recording, slides, and eventually transcripts. Kristin at this time we have no other questions. I wondered do you have closing comments you’d like to make.

Kristin de Groot: Not at this time. Well just one last thing. If you are on a VA research project and are using OMOP, the VA OMOP data right now and you are also using Medicare data and would like to become one of the testers, please let us know. We would like input before the data gets released to the larger research community.

Rob: Well thank you very much for preparing and presenting today. Audience members when I close the webinar momentarily you will be presented with a short survey. Please do take a few moments to provide answers to that survey. We do review and count on your answers to continue to bring you high-quality Cyberseminars such as this one. Once again Kristin de Groot, thank you for preparing and presenting today. And with that I will wish everyone a good day.

[ END OF AUDIO ]