BroadcastMed | Grand Rounds: Measuring the Value of Integrated Care Delivery: Complexities of Identifying Delivery Model Design

JON EBBERT: Good afternoon, everybody. Welcome to CCATS Grand Rounds. I'm Dr. Jon Owen Ebbert, the associate medical director for the Kern Center for the Science of Health Care Delivery. And I'm excited and delighted to introduce Dr. Sarah Meier, who is a research associate in the Division of Health Care Policy and Research and also in the Kern Center with us.

Prior to joining the Mayo Clinic, she was a program analyst in the office of the Assistant Secretary for Planning and Evaluation at the US Department of Health and Human Services. And we were just chatting about that briefly. While it's fun to rub elbows with the power brokers, it's stressful, right? Did you get an ulcer? Was it--

SARAH MEIER: No.

JON EBBERT: Wasn't? OK. It wasn't officially diagnosed, anyway. OK. So, her background prior to-- her background prior to that included evaluation of Medicaid coverage expansion in the state of Wisconsin. And then in the US Department of Health and Human Services, she worked on the implementation of the Medicaid specific provisions of the Affordable Care Act. And some of those provisions changed while she was working on the implementation.

Her current research focuses on risk adjustment in managed competition, development of medical risk measures and poverty assessment, and evaluation of delivery model design. She holds a PhD in Population Health from the University of Wisconsin-Madison and a Masters of Science in Health, Economics, Policy, and Law from Erasmus University in the Netherlands.

Today, her talk is entitled "Measuring the Value of Integrated Care Delivery: Complexities of Identifying Delivery Model Design." So, Dr. Sarah Meier, thanks so much.

SARAH MEIER: Thank you. OK, so, I just want to start out today by acknowledging my research team, particularly my two analysts, Jonathan and Christie, who, as we move through this presentation, you'll see how extensive the data work they've been doing is and really how complicated it is.

I'll start out by saying I have no disclosures and the planning committee also has no disclosures. The learning objectives for the presentation today-- I am going to actually spend about the first third of the presentation talking about the policy context surrounding this research problem. So, when you leave this presentation, you should be able to identify recent policy initiatives designed to reduce fragmentation in health care service delivery and promote high value care.

After I get through about this third of the presentation and contextualize our work, I will be spending a lot of time talking about our data work and the data sets that we're working with. Largely because this project is quite large, we're conducting it in multiple stages. And really the first stage of our work has been getting to know the data sources.

So, as I walk you through that, you should be able to understand and when we're done discuss the data limitations and complexities of large-scale delivery model research. And then finally, specific to some of the early stage findings that I'm going to show you and some of the decisions we've had to make as we adapted our project when we got to our data sources, when we're done with this presentation, you should be able to describe why integration and health care research remains a broadly defined term.

So my outline for today-- I'll start out by providing our high-level research objectives. As I mentioned, I'll move to a discussion of the policy context. I will spend a fair amount of time describing our data sources, particularly a very unique data source we have to look at delivery model structure across the United States. I will outline our research strategy, and then I'll conclude with some early-stage results and the implications of our findings.

So our research objectives. Aim one, where we are going to be spending most of our time, is to determine the association between receiving care in an integrated delivery versus other settings and utilization costs and quality outcomes. So, partially for data reasons, we decided to focus specifically on fee-for-service Medicare beneficiaries with multiple chronic conditions. And, in addition to high-level delivery model analysis, we're also interested in whether there's a relationship between additional provider and setting characteristics and our cost, quality, and utilization outcomes.

And then aim two. When we've gotten through with aim one, we are going to be examining the relationship between care model transitions and health utilization, cost, and quality outcomes for fee-for-service beneficiaries who receive some care in an integrated setting. So we're actually going to be working with a panel data design with multiple years of data. And, as we all know, people don't always seek care in the same setting, so the second stage will attempt to look at how these transitions in and out of different model designs impact outcome.

I am going to start out with the policy context. As I'm sure we're all aware, the Patient Protection and Affordable Care Act, signed into law on March 23, 2010, is really dramatically transforming the United States health care system. So probably most people in the room are aware that the ACA transforms health care coverage and access.

This is really where a lot of the news focus has been, where a lot of the political debates surrounding the ACA have been, and where we've probably seen the most dramatic shifts immediately in the early years following the rollout.

So, what I think people are less aware of, although here in a clinical setting you may be starting to understand or see the second phase of transformation, which is that the ACA is also dramatically transforming health-care payments and delivery. So I will get into specifics of what that looks like in the next couple of slides. But I'll just briefly highlight some of the major health care access and coverage transformations.

So, the first one is Medicaid expansion, one of the major ways of facilitating access, as we all know. Following the Supreme Court ruling, that has become optional. The individual mandate, also a component of the concept of how access could be expanded and the insurance market would not destabilize. In addition to those two pieces, which are the two pieces that, again, probably receive the most news coverage, there has really been a dramatic shift in how the government regulates insurance markets.

So, in this field, this is really referred to as the three Rs-- reinsurance, risk adjustment, and risk corridors. This is really meant to level the playing field among insurers so that payment is fair, irrespective of the risk profile of enrollees and different insurers plans. Risk adjustment actually stays-- it stays for good.

The other two-- reinsurance and risk corridors-- sunset. They're only really there to stabilize the market in the early years when there's a lot of uncertainty about how the new reforms are going to work. At the same time, insurance market reforms-- they're broader. We're also seeing the insurance marketplaces-- the state-based and federally facilitated marketplaces-- premium subsidies and cost-sharing subsidies acting as a component of enabling access to coverage within those marketplaces.

And then, finally, this last one I highlighted because everyone in the room actually should be aware of this basic health plan. Minnesota is the first state to move forward with this this year. And it's also another provision that is probably a little less known but, sitting here in Minnesota, we should be aware of it.

Accessing coverage. Hopefully you've heard of some of those changes. At the same time and probably a little slower starting and picking up some speed, there are a large number of payment reforms. So, the common phrase that you probably hear is that we are working on moving the health care system from payment for volume to payment for value. So, as you're probably aware, historically Medicare is largely fee-for-service payment design.

With the introduction of Medicare Advantage, there's a fair amount of capitation. We've kind of seen the same trends in Medicaid. But broadly speaking, these payment reform initiatives are really supposed to move away from fee-for-service that is not at all connected to quality and outcomes to either global payments that's up front or some form of fee for service that really accounts for whether the outcomes received are really the outcomes we were looking for.

So examples of these programs include the Hospital Readmissions Reduction program, the Hospital Acquired Conditions Reduction program, Hospital Value-Based Purchasing, which I think has received a little more notice in the news and really in the health services research literature. Moving forward right now is Physician Value-Based Payments. And then the Bundled Payments for Care Improvement Initiative.

And then beyond these initiatives, there's other ways that payment is actually being reduced. Examples include a reduction in the growth of payment rates in Medicare. And then also there's kind of been a longstanding debate about whether Medicare Advantage plans are overpaid for the Medicare enrollees who choose to select into those plans. And currently payments to those plans are being reduced.

OK, so a lot of work really to kind of either try to take money out of the system or at least to diminish the growth in how much money we spend on health care system over time. The ACA has not only been focused on payment but also restructuring the models of care that are behind payment.

So, the two ways that this has really been done or at least the most focus has been placed in these areas is to promote the development of the ACOs-- accountable care organizations-- which are groups of providers and organizations that are accountable for a population and the total cost of care for that population.

These types of programs are supposed to really influence collaboration across providers and entities that maybe typically don't always work together. Sometimes that is not just across clinical providers but could be across social support systems. Some of the Medicaid-oriented ACOs are a little more focused in that area.

So, the top four lines here are four ACO programs. You've probably heard about the Medicare Shared Savings Program. I think there's some recent literature coming out starting to evaluate that. The Pioneer ACO Program, which really was initially envisioned for kind of more advanced entities, perhaps like a Mayo Clinic type, who really have experience integrating and coordinating care.

The next generation ACOs, I believe, just started this year or is starting as we move forward. In addition to ACOs, a focus on the development of medical homes, which have a primary care orientation. And I just want to highlight that through all this, all of these new models, nobody really knows. There's a lot of hypotheses, right, that they're going to work. Because otherwise we wouldn't be doing them.

But the actual research and evaluation that says, OK, these models of care delivery, they improve quality or, at a minimum, they hold quality constant and reduce cost, the research behind that is actually quite limited. So what CMS does, when they're promoting these programs and approving them, is they attach kind of an evaluation requirement to it.

So what we're seeing across the US-- and the next slide will get into this-- is a large number of demonstration programs that are approved for a certain amount of time. They have an evaluation component attached to them. And, at the end of the day, ideally we find out whether they were or weren't effective. And at that point, I think part of what the ACA did, it really actually gave the secretary some power to expand those programs that were proving to be effective.

The last two sides really focused on how the federal government is really looking to kind of pull money out of the system and at the same time transform the system. And the good news is that there really have been-- there has been a sincere effort to try to partner and work with providers and states as they're doing these transformations and to try to fund the transformation.

So what we've seen at CMS is a kind of a reorganization that includes the development of new centers and initiatives. And as I speak about this demonstration requirement, these centers have played a key role in funding the demonstrations and improving. So the Center for Medicare and Medicaid Innovation-- CMMI or also known as the Innovation Center-- is probably, really, the headline entity within CMS that is running these programs.

The programs that have been approved through CMMI currently impact 50 states, 60,000 providers, and 2.5 million patients. So it's a huge reach. There really is not a single state that has not been impacted by this concept of delivery model transformation. Within CMMI, we have the State Innovation Models Initiative, which has been responsible for awarding SIM grants. 39 states have received SIM grants. Minnesota is one of them. Wisconsin is one of them.

And where this differs from maybe some of the other programs that are more broadly under CMMI is that State Innovation Models are typically-- the state is actually working with private payers to transform kind of the component of the delivery system that the state would have control over. The Medicaid Innovation Accelerator Program is also within CMMI. That's specifically focused on Medicaid.

The Medicare-Medicaid Coordination Office-- also known as the Duals Office-- is focused on the dual eligible population. So these are individuals who are dually eligible for Medicare and Medicaid. They are disproportionately expensive in both programs. There are a lot of issues with the coordination of their care, the structure of payments for them. The Duals Office is really tasked with trying to resolve some of these issues.

And then, finally, if you haven't heard about CMMI and you haven't heard about the Duals Office, hopefully you've heard about PCORI, the Patient-Centered Outcomes Research Institute. PCORI has funded 365 projects in 39 states. I'm pretty sure some of our researchers here have PCORI grants. And that is really looking at comparative effectiveness research with a focus on pulling the patient into the research experience and the research design.

So hopefully I've convinced you that we are really in a historic moment in health care in the United States. It's not just because of changes to coverage and access, but it's also because kind of the old business as usual, in terms of how payment and how care delivery works in the health care system, is changing. And it's just not possible to hide in the corner from it.

So, with this change in mind, one of the things that's really important, as I've highlighted, is that everybody needs to be evaluating what they're doing and showing evidence for the effectiveness of their models. So, we here were interested in the question of, well, an entity like Mayo Clinic, who has a strong model of integrated care delivery. It's not necessarily an ACO, so we're not falling under these ACO demonstrations.

What's the effectiveness of our care? We were lucky enough to have a really cool data source-- the IMS Health Care Organization Services database that is essentially-- it's a quasi universe of providers in the United States, specifically DOs and MDs. And it is also a quasi universe of health care businesses in the United States.

So we decided, hey, we want to take this database, we want to look at the structure of care delivery across all states, and we're going to see if we can understand how receiving access in an IDN-- or an integrated delivery network-- how that impacts outcomes. To do that, as I mentioned previously, we decided to use Medicare data-- claims data from Medicare you can obtain research access.

So we are actually going to have access-- or we do have access to four years of data for a longitudinal cohort. One of the additional initiatives that CMS has started is the ability to access this data through the Virtual Research Data Center. So we're also actually kind of one of the first teams outside of the federal government that is working with the data in this capacity. And we are linking these two data sources to examine our research questions.

So, before I get into the details of the data, which we'll kind of back up from this specific focus, I just want to give you guys these definitions, so you understand what we're working with. The HCOS database identifies integrated delivery networks. It allows us to link every single business that is in the database to a network, if it belongs in that network. And the networks-- so it identifies IDNs and it also identifies a different type of network.

So an IDN network within the database is a type of corporate parent. It's an organization that has direct responsibility for centralizing the purchasing or contracting of its affiliated hospitals and ancillary care facilities. And then the key point is that it offers a continuum of care through services at acute and non-acute sites.

In contrast, we can also identify networks in this database that are what are referred to as corporate owner networks. They're very similar, but the key point is that they do not offer a continuum of health care. So we can structure networks for every major network. We can tie physicians to these networks. And then, in addition, kind of the third category that is not on this slide is of course a practice, a hospital. It can be independent.

OK. So, keep those definitions in the back of your mind. The last section of this presentation is really going to focus on what that IDN definition does and doesn't mean and what some of the complexities are working with that definition. So, backing up to a broader focus on these data, the IMS HCOS database-- it's a relational database. It has detailed provider and business data.

And our data actually reflects a snapshot of the United States system. So we have an August 2013 snapshot. That's the data we have. It doesn't perfectly align with our years of Medicare data access. But that's kind of-- you're never going to have the perfect data and the perfect data set. So that's kind of where we're standing.

The key components of this database include a business file. So this is kind of this quasi universe of health care businesses in the United States. It includes relational files. So this is ownership and contracting relationships, kind of parent-child-grandchild relationships, and even things like academic relationships.

We have supplementary business data, so all sorts of fascinating information, like bed type, facts about the beds, financial information to the extent that it was reported to IMS, and all sorts of additional details. The database also includes clinical data and then a provider file and provider affiliation data.

So the benefits of working with these data for us is the ability to build networks and hierarchies among businesses and professionals, the availability of professional affiliations and profile information, the availability of class of trade information-- standardized class of trade information-- which includes businesses that are really assigned to a classification and then a facility type and a specialty. And I'll get into those specifics as we move ahead.

But solid definitions that really help us understand kind of where these businesses fit in in the delivery system. And then, as I mentioned, we can identify IDNs, non-IDN type networks-- what I didn't mention is that there's owner subsidiaries also included under the parent umbrella, which is going to turn out to be a key piece of how we're approaching this project-- and then independent businesses.

So the business file. Organizations are at different levels. We kind of start with this big file. We merge in relational information, and we use that to construct networks. Then we kind of go over to the provider file and we merge that on to all of this business information, and we link providers.

So, because these data really fall at different levels, one of the key things that we kind of realized is a really key component of structure and analysis-- that the data is deciding what unit you're going to focus on, whether it's a delivery setting-delivery site unit, whether it's an owner subsidiary unit, or whether it's really the corporate top. And you will start to see why this is important. And then, again, as I've mentioned, the network mapping is really a key component of these data.

So, further expanding on kind of these basic-- this is a certain type of business; we can define it; we can understand who it's connected to-- it's actually quite astounding the level of detail we have about each of these organizations. We have their name and address. We have things like their case mix index, total licensed beds, census beds, staff beds, number of surgeries if it's a business type that would conduct surgeries.

And then we actually have some really interesting variables, things like whether the business operates with a formulary, whether there's EMRS present, whether the business is participating in some form of a pay-for-performance initiative, and other relevant variables of that sort.

And then I just kind of am highlighting here some of the additional details of how it's not just that we can identify a bed count, but we can identify types of beds. It's not just that there's very basic financial information, but there's kind of many facets of financial information when it's reported.

So, to further refine this kind of organization class of trade information that we have available so you can understand this, as I've mentioned, each organization is defined with a classification, a facility type, a specialty, if applicable. The kind of top of the network type organizations are going to be an IDN or corporate owner. Under that is going to be an owner subsidiary, and then beneath that will be delivery site.

This slide here and the next one actually gives you kind of a view of how far-reaching these data are. So bolded is-- when I say that we know a business's classification, the bolded categories are classifications. The next line item is facility type. And then it would be too much to show on this slide, but some of these facility types have specialties.

So, again, to kind of show the breadth of the data, we can identify all academic medical schools, we can identify all hospitals, understand the type of hospital. There are a large number of outpatient center types. And then it's not just kind of the traditional-- so, when we think of health care delivery we think of hospitals and outpatient centers.

But there's all other sorts of places where care might be delivered. One would be a correctional facility. Another would be that laboratories play a key role in the care process. Here you can see we have pharmacy information. We have information on research institutes, information on the long-term care component of the marketplace. So really, really extensive data.

And then the professional file. So we have provider-level data. It includes numerous provider types. So, for the work that we're doing here, we decided to focus on MDs and DOs. This is actually-- initially we wanted to incorporate nurse practitioners, physician assistants. We ended up finding out that-- so, one of the questions in the back of your mind is, so where do all these data come from? And the reality is that different components of the database had different levels of data quality.

There's associations, there's reasons why it would not be that hard to come up with a universal list of physicians in the United States. That is more difficult when you're looking for a universal list of nurses. So we kind of reached a point in our discussions with the data vendor that, for our project, where the data quality is falling and what we need to do, we're just going to focus it on MD and DOs.

But I do want to highlight that it's not just MDs, DOs, nurse practitioners, physician assistants. We have vets, pharmacists, I believe dentists. It was really actually quite surprising, the number of provider types that were included.

So, for each of these providers, we can understand whether they're operating in a provider role, an administrative contact role, or both. We had key identifiers that allow us to link over to the CMS claims data. So the NPI, UPIN, DEA numbers. We have information on primary, secondary, and tertiary specialty. And then we can also see whether a provider is active or semi-retired.

So I've spent a lot of time talking about how this is in a relational database, how the database falls at different levels. Hopefully you've come to understand that what we actually have and what the analysts have to work with are multiple files that we have to figure out how to link across. At the end of the day, what we're able to do, for every major health system in the United States, we can build out a network.

And this is kind of what a hypothetical network structure looks like. So at the top, we will have a corporate parent. That's either an IDN or a non-IDN type parent. Beneath that we might have corporate owners. We might not. Or, sorry-- we might have owner subsidiaries, we might not. Beneath that level, we kind of have the delivery setting site level. So here, I focused on hospital types and outpatient center types. And then, below that level we have the actual clinicians.

And a clinician can have more than one affiliation. They can be affiliated to multiple sites within a network. They can be affiliated across networks. And some of the clinicians-- we're not just talking about one or two affiliations, we're talking about-- I can't remember the max, but it's a lot.

And, then, just a brief discussion of the Medicare data. And so, as I mentioned, we are focusing on fee-for-service Medicare beneficiaries. Our actual analysis years are going to be 2009, '10, '11, and '12. So four years of panel data.

We do actually have access to 2008 for some of the data files, because ultimately we're going to need to attribute patients to providers and providers to delivery settings. And some of that attribution work, for the first year we would need prior year data. So we also requested access to that data.

To be included in our cohort, an individual needed to have at least one year of multiple chronic conditions status. We define that as two or more non-cancer Chronic Condition Warehouse conditions. They needed to have continuous part A and B enrollment during the same year, no end-stage renal disease. And then they also had to have Medicare as primary payer.

Just so you can understand, again, the breadth of data files that we have access to, for this cohort we have access to kind of the summary files, which are really nice. They kind of roll up a lot of the utilization and clinical data into some key variables that researchers might tend to incorporate into their study. So, that includes the costs and utilization segment file. This is things like hospitalization counts, emergency visit counts, summary cost data.

On the chronic conditions segment, which is really a set of predefined chronic conditions that we can see whether an enrollee has ever had this specific condition or whether they had it in the year of focus. And then the enrollment file, which is kind of like a base file that helps us understand category of eligibility. It has some basic demographic information-- age and gender-- some very limited socio-economic information but that's still useful.

So that-- the summary files. In addition to that, the MedPAR Carrier, DME, HHA, hospital, and outpatient research files. And then, finally, for a subset of our cohort, we're going to have access to part D Drug Event information. So this is actually what our full cohort looks like.

Each column is the count of beneficiaries who meet the criteria in one year. Final column is the number who continuously meet that criteria over four years. And you can see that, when we've applied all of our criteria, at the end of the day we are going to be working with about 20 million beneficiaries, which is really a huge number. And ideally we'll be able to link this over to all this delivery model data.

So the research strategy. As you can see, this is a lot of data. We It took us, actually, quite a bit of time to gain access to the Medicare data. We've really had a strategy of working on this project that has been in multiple phases. The first phase has been really starting to understand the IMS data. So this has involved classification of organizations and providers and organization relationships.

One of the first things we did-- and when you're working with data like this it's really important that you do this type of thing-- is we built data dictionaries. We had our baseline definitions and understanding of what should be in the data, but things we needed to understand is how often does a variable have a missing value? So we built those data dictionaries for the professional and the business files.

We then went on to do basic descriptive analysis, looking at the distributions of IDNs across the US, looking at the distribution of independent businesses and non-IDN networks. Also, as part of this descriptive analysis, we actually sat down and built the network. Like that generic slide. So not all of them. But we built the Mayo network, because that's the network we know a lot about. And then we also decided to build the Kaiser network as a second way of looking at these data.

And, again, that was really a process of understanding how all these data pieces fit together and throughout the process kind of brought up a lot of questions about what does this mean, why is this missing, and was a really informative piece of work to get to where we are now.

We've had access to the Medicare data maybe a couple of months. So we've just about finished up the baseline descriptive analysis of these files. I showed you very basic descriptive information, but we went a lot further than that. And then the end of this stage has been-- and this is what the end of the presentation will focus on-- realizing that the IDN definition that we have in the data is really not going to be sufficient for what we want to end up looking at.

IDN is a very broadly defined term. And we wanted to see if we could come up with a way to further refine it and to kind of create a spectrum of integration or subclassifications of integration for purposes of incorporating into our analysis. To do that, we have been doing cluster analysis work, which we are wrapping up right now.

We are currently moving into stage 2, which is finishing this classification work, finishing these clusters, and then we're going to have to begin the process of attribution. And when we are completely done with these two stages, our goal is to have an analytic file with four years of longitudinal data, a panel data set, basically, that will include static and time-varying beneficiary level information.

So demographic information, SES information, HCC risk score at the annual level, delivery model information that's basically been pulled in from our analysis of the IMS database and attributed and assigned at the annual level, cost utilization and quality outcome measures, which we are planning on pulling from a lot of work that's already been done out in the research literature, Choosing Wisely measures, Hospital Compare measures.

We will be kind of implementing those in the Medicare data and then linking them into our data set. And, at the end of the day, once we have this data set, the process of actually kind of developing the econometric models to examine the relationship between care integration and outcome is actually the easy part. So that's the final stage.

OK. So, now I'm just going to get into the final section, where you can see some of the basic results from what we've been doing with the descriptive work. I know this is a little hard to read, but hopefully this will give you an understanding of not only how complex the data are but how complex the US health care system is.

I told you we built out the Mayo network and the Kaiser network. This is what the Mayo network looks like. And if you know the history of the clinic and the health system, it's really cool because you can start to see that in these data.

At the top, Mayo Foundation is our corporate parent. It's an IDN in the database. I told you we have owner subsidiaries, which-- sorry you can't read it, but what you'll start to see is the owner subsidiaries include our Arizona and our Florida locations. They include places like La Crosse, Eau Claire. They include the general health system and then a couple of entities that are actually broken out on their own.

So that's the second level. Rochester, we figured out, is actually the Mayo Foundation level. So you will see the arrows going over to the side here and the boxes that are connected on the left that directly link to Mayo Foundation includes Rochester. Below that then, the actual delivery sites that I've been speaking about.

So this summarizes kind of the hospital outpatient centers, the types of outpatient centers. The clinician information is not linked to this. But then, finally, academic relationships are also structured in. The light blue are actually academic relationships that exist with the Mayo network that are non-ownership based. Red lines are academic relationships. Blue lines are ownership-oriented type relationships.

That was the Mayo network. I also told you that we did descriptive look-- or descriptive analysis, looking at, basically, the distribution of IDNs across the United States and how specific site-level organizations linked to IDNs. So, what you're seeing here is a map of the United States. These are hospital referral regions. The boxes are quartiles. And you are seeing the percent of hospitals in each HRR that are connected to an IDN as opposed to being, for example, in a non-IDN type network.

This next slide-- one of the outpatient center types is a single specialty medical group. This is the proportion of single specialty medical groups in each HRR that are linked to an IDN. And the nice thing is, what you're seeing-- for example, the Midwest is known for being a more integrated kind of health care delivery system compared to other areas of the country. And you're seeing that show up with the upper Midwest.

And then the last mapping slide. Each medical group that-- former slides showed medical groups in general, which we can actually break down into is it a cardiology type medical group? Is it some other type medical group? One of the breakdowns is actually multi-specialty group practice. And this is showing that same type of analysis from multi-specialty group practices.

That's kind of how far spread IDNs are and how much kind of control per se they have over the delivery system in different areas of the US. So now I'm going to get into a more specific discussion of what this IDN term means. We know that an IDN, there is kind of an ownership structure where the delivery sites are in some way, shape, or form owned by the head of the network.

And we also know that there's a continuum of services. So what does that exactly mean? And are all IDNs in the data set the same? We realized pretty early on that not every single IDN in the data set is the same. Within Minnesota, if we were just to focus on the corporate level, which is kind of that Mayo Foundation level, we had 29 IDNs.

An IDN in Minnesota can have anywhere from nine to just short of 13,000 providers connected to it. And it can have anywhere from 1 to 27 hospitals connected to it. So if we were simply to do an analysis where we looked at outcomes associated with receiving care in an IDN, you can see that we have a problem that this is not-- this is really lumping together a lot of networks that are quite different.

Then, if we focused on the corporate owner plus owner subsidiary level, this number-- these numbers changed to there being really 45 IDN-esque kind of smaller networks in Minnesota. The number of attributed physicians moves from 9 to 8,000. And the number of hospitals attributed moves from 0 to 17.

One thing I want to say about the data is the reason you're seeing the difference in the numbers for attributed physicians and attributed hospitals between these two levels of the data is that, when we're looking at the corporate level, that count attributes to Mayo Foundation hospitals that are outside of the state of Minnesota. So Arizona, Florida, Wisconsin, they're captured in that count, even though I'm looking at Minnesota, because Mayo Foundation is in Minnesota.

A similar thing is occurring with physicians, which is why, when we drop down to this owner subsidiary level, owner subsidiaries are locally identified in the data, and then we can focus only on the IDN owner subsidiaries that are specifically in Minnesota. One other thing I want to highlight is that these data also include the Indian Health Service and the Veterans Health Administration as networks that we can see in our data.

We decided that IDN is far reaching. We need to come up with a way to kind of try to distinguish across types of IDNs. And so we decided to do cluster analysis work, where we took a large number of the attributes that we have in our database and we tried to see how similar different IDNs were to each other, based on the values filled in for these variables.

So, to do this, we did this across the United States for every single network. But, to look at the data, because it's a lot of information, we focused on Minnesota and we focused on Mayo Clinic specifically. This slide here is just providing your base summary of who's in the Mayo network, who the owner subsidiaries are.

This next slide is talking about this approach that we're going to be taking to doing the clustering analysis. So what we did is we compiled two new data sets from this database. We compiled one at the network level, which is the Mayo Foundation level, one at the owner subsidiary level, which would treat each owner subsidiary within the Mayo Foundation separately.

And we came up with a large list of characteristics from our data set to kind of define each of these little baby networks. So this included the number of business type at the level of specialty, the number of physicians at the level of specialty, data on ownership, management, and leasing structure, financial data, case mix and discharge data, and pretty much anything that we could find in the data set that we were able to incorporate into this.

I want to mention that this included somewhere from about 100-- well, around 150 variables for each network. We then went on to do K means cluster analysis where we had to specify the number of clusters that we wanted the IDNs to fall into. So on the next slide what I'm going to show you is the results of saying we want to create 3, 4, 6, and 8 clusters. We also looked at 10 clusters.

So I mentioned we had this data set of about 100-- well, at the end of the day it was fewer than 150 variables-- but a large number of variables. And once we had the status there, we had to kind of pare it down. And there were some easy and hard ways to do that. In the first place, some of our variables the data was missing in too many places to use that in the cluster analysis.

In other situations, there were variables that we really wanted to use because we thought they got at the characteristics of care delivery, for example, whether EMRs are present. But there was limited variability in the variable across networks, which made it a problem to incorporate it in the clustering work.

Collinearity was another issue that led us to drop variables. Those things kind of resulted in dropping variables, making decisions to summarize and roll up variables. And then the last pieces that, as we were doing the analysis, there were statistical reasons why some of the variables needed to be transformed.

What you're seeing here is the list of variables we included. So some summary measures on physician counts, the characteristics of how physicians are affiliated within a network. So whether they're just affiliated with a hospital or whether they're affiliated with two or three hospitals and an outpatient center.

Let's see-- we also looked at things like whether every entity in a network is in the same HRR, whether every entity in the network is in the same state, what the ownership structure across the network is. And, then we had information on discharges, so we looked at percent Medicaid and Medicare discharges.

What this slide shows is it rolls up the results. And, like I said, we did this for all of the United States. But here I'm focusing on the results for Minnesota. So, let's see-- I think we had 45 IDN owner subsidiaries in Minnesota. And you're seeing how-- the bold is showing how they fall across the clusters. So each cluster, conceptually, should be grouping together IDNs that are similar on the variables that we put into the cluster analysis.

And then I'm also pulling out on the slide the Mayo network-specific entities, irrespective of whether they're in Minnesota or not. So three clusters, you're seeing that everybody except Eau Claire and Mayo Foundation are put into the same cluster.

Four clusters, we're seeing the same results. Six clusters, things start to break out more. And the general health system, Eau Claire, and La Crosse are grouped together. Mayo Foundation is on its own. And then the other entities are in another cluster. And then, fairly similar results for the eight cluster approach, where Eau Claire is put off on its own.

And I won't outline specifics of this, but this is for the six cluster solution, and it's showing you how each cluster differs. So some key characteristics of the clusters that were varied were things like the number of physicians in the network. Another one that actually varied a lot is the percent of entities in the same state as the network and the percent of entities in the same HRR. And then also Medicaid discharges and Medicare discharges-- there was a fair amount of variability.

That's kind of where we are on this project. As you can see, it's really extensive data work. We are currently in a phase of finalizing the clustering work. We're starting the process of attribution, linking across data sets. And then, as I've mentioned, final phase will be looking at the relationship between setting and outcome.

The implications, as you should take away from the first section of this presentation, the current policy environment is really focused on transforming the delivery system to realize value. There is a lot of support behind the idea that we need to be creating ACOs, we need to be promoting care coordination, structuring innovative payment models.

But at the same time, the evidence for the effectiveness of integrated care delivery arrangements, ACOs-- all these new models-- is actually quite limited. And we are really-- not just here but across the health system, across academic settings within the government, the contracting organizations that do a lot of contracting work. This is kind of one of the key areas that health services research is focused on right now is evaluating this type of transformation in the health system.

And then I hope you've understood that while this is a really important research question, there's a lot of complexities to answering it. We happen to have access to really phenomenal databases. But even with databases as comprehensive as we have, there are still challenges. You've seen the challenge with defining integration, measuring it.

And then also what you-- I didn't highlight on the slide, but a lot of the components that we have had to work with are structural. They kind of are characteristics of who's in the business and how many of them. They aren't really actually characteristics about the actual process of care that's going on within each of these networks. And that's another area where the data sources just really aren't as strong as we or external researchers would really like them to be.

So, that concludes my presentation and I'm happy to take any questions.