Medicaid Innovation Accelerator Program (IAP)
Webinar on Unique IDs and Medicaid Data
January 25, 2018

Jessie Parker (JP), Analyst in CMCS and Government Task Lead overseeing the IAP Data Analytic Technical Support Work: Our speakers for today are:

- Manjusha Gokhale (MG), Senior Data Scientist, Truven Health Analytics, an IBM Company. She will be providing methodologically-oriented detail view to common challenges and strategies in using beneficiary IDs in Medicaid data, as well as a high-level overview of probabilistic and determinants of matching and data linkages.

- Bruce Greenstein (BG), Chief Technology Officer, U.S. Department of Health and Human Services. He will be talking on this experience and his time as Secretary of the Department of Health and Hospitals in Louisiana and addressing the challenges of Medicaid eligibility data from a programmatic perspective.

A brief summary of our agenda:

- Overview of the Medicaid Innovation Accelerator Program (IAP)

- Working with Beneficiary Identifiers (IDs): Manju

- Linkage Across Data Sources: Manju

- National Death Index: Manju

- Challenges of Medicaid Eligibility Data: Bruce

- Practical Applications for Linking to Vital Statistics: Bruce

- Joint Question & Answer Session

- Takeaways from Today's Webinar

This seminar is produced through the Medicaid IAP. IAP is a cross-center collaboration of CMS intended to support states' payment and delivery system reform efforts. We are located in the data analytics functional area and one of the approaches we are using to increase data analytic capacity with in-state agencies is to host webinars such as this one on data-related topics and challenges that are relevant across a wide area of states.

Our primary goals for today's webinars are:

- To walk through common challenges in working with Medicaid enrollment data and present solutions or best practices to those challenges.

- We will also cover the pros and cons of probabilistic and determinants in matching.

- Walk through how to link to the National Death Index (NDI) and other linkages.

- Address difficulties in working with Medicaid enrollment data from the programmatic perspective.

Our goal is that you will walk away from today with specific approaches that you can apply in your own local data environment. Our first speaker will be Manju Gokhale from Truven Health Analytics who will discuss how to approach these questions from her methodological perspective.

MG: I'd like to talk you through challenges in Medicaid beneficiary data, or any kind of enrollment data, and some strategies to link them to other data sources. One of the first things we'd like to look at is thinking about beneficiary IDs in Medicaid data. It's important to know who your beneficiary is to pay a claim. But we're also interested in having an accurate assessment of the beneficiary-- who the person is-- in order to do longitudinal analyses, which require the correct denominator, such as comparison to commonly available benchmarks like the ARC indicators, assessment of expenditures, PMPM analysis, and any kind of population health analysis.

Typically, Medicaid enrollee identifiers are assigned by each state Medicaid agency. There's not one best way to assign a Medicaid agency or one standard they all follow, although people like me who work with this kind of data wish that was the case, but it's not. Usually it's identified during the enrollment along with other highly identifiable information about the person, such as Social Security number, date of birth, first name, last name, gender, address. Besides the first two, these last four can all change over time, and that's one of the issues we have when we're trying to look at enrollees across time. People have a tendency to move or get married or other things that happen. In an ideal world of health infomatics, you could just count the number of unique Medicaid enrollees in a year. But if you did do that, you'd get a different number than the total number of Medicaid enrollees, and I'll outline what some of those issues are.

Issues with enrollment include carve-outs for managed care, behavioral health, pharmacy coverage. You can have a single person, but they might have maybe three different IDs based on what the state has put into play. They might not know that, but if you simply added up the number of Medicaid enrollees, you'd get a number that was three times the true number of Medicaid enrollees.

Another issue is, and this comes into play in linkage, is that during birth mother and baby claims are often linked together, especially for well babies who might not generate their own medical plan. That is an issue. Then the Medicaid population has a lot of enrollment and disenrollment and that's an issue throughout the year.

Let me go into these in more detail. In specialty carve-outs are arrangements, where let's say the state has contracted a third-party entity to administer the care given for certain services, the issue is you have multiple enrollee identifiers, one for medical, one for behavioral health. We would recommend in a case like this that you maintain a crosswalk of specialty carve-out identifiers to state Medicaid enrollee identifies. So if you have an organization that is going to administer your state behavioral health and they're going to use their own identification system, maintain a crosswalk that has the two so you can look at the person's full health experience.

Another type of carve-out is a vertical carve-out, which is a medical managed care plan. The issue is that the individual is listed in the Medicaid enrollment but they could also be assigned another internal enrollment identifier by the health plan. So this issue is a case where you've got all the information and enrollment data but you don't have any of the claims in your Medicaid claims. They seem like great people to have because they don't have any utilization and they're saving a lot of money, but again we would recommend that you maintain a crosswalk, a vertical carve-out, enrollee identifier to state Medicaid

enrollee identifiers, and that when you do any kind of cost or use analysis with a Medicaid claim, that you exclude these individuals. Alternatively, you could bring in their utilization experience but this is the simplest way, and you could apply the same sort of logic to other cases.

The third thing I'd like to talk about is the combined mother and baby enrollment. Healthy babies are usually not enrolled in Medicare exactly at the time of birth. So some current enrollment methods undercount healthy babies in Medicaid enrollment. So we would recommend confirming the number of infant enrollees by augmenting figures with information from birth records, vital statistics, as well as hospital discharge claims.

The final enrollee identifier issue is disenrollment and reenrollment. Some individuals will dis-enroll from Medicaid and then reenroll and get assigned a different Medicaid ID. Hopefully this is happening less but this will certainly happen in states such as New England where people might go from living in Massachusetts to living in New Hampshire to living in Vermont, all very close to each other, but they may get three different Medicaid IDs. The person might also remain in the same place but get married and change their name and somebody might not be aware of that, somebody who was not on Medicaid before.

We would really highly recommend you use something like the Social Security number to confirm that the individual does not have a prior Medicaid enroll ID or some information about that person like name, address, maybe even questions.

We recommend that states maintain a Master Patient Index, a method of aggregating the information from disparate data sources. An MPI should contain only those fields which uniquely identify a person and usually do not change, such as Medicaid ID, Social Security number, date of birth, gender. Ideally, Medicaid enrollee information should be consolidated into an MPI and the MPI in an ideal case should equal the number of unique Medicaid enrollees that there are.

Now I'm going to jump into determinants of probabilistic matching. Why I'm moving into this field is that oftentimes people build their MPI based on enrollment data, and enrollment data typically has let's say if somebody was enrolled for 12 months they would have 12 records for that person. What you would like to do is to consolidate all the information for that person so they only count as one person instead of 12 people.

The way you do that is using two techniques. One is called deterministic matching, and that's what we think of as matching. If somebody has exactly the same name or same address, any kind of merge or exact match on something, everything is the same, that's what a deterministic match is. A probabilistic match uses a statistical approach and calculates the likelihood of a match. If you heard of something called fuzzy matching, that's what probabilistic matching is. It's kind of like things you would know without even thinking about it. Like if you need somebody named Margaret and they also go by Peggy, you might know that's the same person because that's a short form. But computers are not quite as good as the human brain. So we have to build in statistical techniques to approximate thinking the kind of thinking you do without even thinking about it.

In the probabilistic matching, there's a few different techniques people use:

- One is phonetics. Like if they see a name Katherine with a K versus Catherine with a C, that's probably the same person.

- Synonyms: Andy, Andrew, Andie, George and Jorge, 1st and First, Jr. and Junior, those are probably the same.

- Abbreviations: J.D. Salinger.

- Concatenation.

- Misalignment: Robert John might be equal to John Robert. There's some issues with that because those could be two different people.

- Distance: For example, Edit Distance = 876-5309 is probably the same as 876-5390 with a typo.

- Transliteration. Same as Distance.

- Date Similarity. Same as Distance.

- Proximity.

- Noise words: IBM Company is the same as IBM.

Those are some of the techniques people use when they talk about probabilistic linkage. Basically they're just trying to mimic what you can figure out without thinking on a great number of records in the data. To see how likely this is, people will assign weights to how much they trust this method. For example, Katherine versus Catherine, that's likely the same person so it will have a higher weight, or the synonyms. But misalignment maybe will have a lower weight and we'll go into a small example.

The advantages of deterministic matching, which we think of as matching, are:

- Confidence of match.

- Easy to understand and explain. I look for people with the same name and same birthdate and they matched in all the fields.

- Not dependent on knowledge of data file. You're just matching these two files together.

- Can use all matching fields and then drop criteria one-by-one to capture remaining non-matches. Say you have to match on five out of five, so it's first name, last name, date of birth, Social Security number, and address or city. For everything that doesn't match you can drop it to they have to match on four of these fields, etc. After review you can capture the remaining non-matches.

The disadvantages are:

- Rigid in structure.

- May undercount denominator. Let's say somebody sometimes spells their address Street and sometimes spells it St. It's something you know but the computer program does not know it.

- Can exclude common errors such as people may move every year or might have roommates or a lot of other things.

Those are problems with deterministic matching.

Probabilistic matching as we were talking before is where we can match across fields with transcriptions, multiple spellings, address changes allows you to maintain a longer longitudinal enrollment file, which is really good for analytics population analysis. But the difficulty is it's hard to describe. There's this one method of how different is one character string than another called the Winkler Method. That's not really easy to describe it to people. It's also hard to do in terms of it takes a lot of time to review, or you can buy some off-the-shelf things that are expensive to implement. You also might have many false matches or non-matches. You might be linking John Francis to Francis John. And you're highly dependent on patterns in your database. Let's say you know that people often flip date of birth, the year and month. You have to know that about your underlying data. If you don't know that, your faith in your probabilistic linkage is probably not as good.

Now I'm going into a quick example of a sample enrollment file. As you can see this is many months of data with one Social Security number and two IDs. You can see it's for Anita Chen, born on 07/07/2007. She lives on 5 Lexington Ave. in Watertown, Massachusetts. Every month she's enrolled in medical care and she has a specialty pharmacy card out. As you can see they have different enrollment Medicaid IDs. If we just added up the number of unique people, we would get a different number of people. You can see it's Anita Chen, Anita Chen, Anita Chen, Anita Chen, but here all of a sudden it changes to Anita Hines but it's the same address, same date of birth, same Social Security number. Next we see Anita Hines move to 6 Watertown Ave. in Lexington, Massachusetts whereas before she was at 5 Lexington Ave., Watertown, Massachusetts. Is this a typo? Did she move? It depends on how much you trust your data. Finally we have an Anita Heinz now living at 6 Watertown Ave. in Lexington, Massachusetts and enrolled in a different plan. This is the kind of data you're going to have to deal with and I'm sure you do encounter that makes up your MPI.

If we use just determinants of matching on Social Security number, luckily because we have Social Security number and it was completely the same there, we would end up with a single person, Anita Heinz. Typically people use the last name listed as the name with the idea that whatever's more current is probably more accurate than earlier fields. But if we didn't have a Social Security number or date of birth and we matched on last name, first name, and address with deterministic matching, we would have ended up with three people: Anita Chen, Anita Hines, and Anita Heinz. Probabilistic linkage would allow us to have one person, Anita Heinz, even without Social Security number or date of birth.

Now I'd like to ask a question: Has your agency used any of the following when working with beneficiary IDs? Have you worked with: Probabilistic matching? Deterministic matching? Both probabilistic and deterministic? Or never used either approach? Can people on the WebX submit their responses? As you can see, from the results people seem to have a lot of experience with both probabilistic and deterministic matching. Wow! That's great. So 45% of you have used both and 8% have used just deterministic.

To on with recommendations, we'd like to:

- Establish a hierarchy of linkage.

- Examine the matches for confirmation and examine a set of non-matches to view patterns in errors.

- After examination, loosen the match criteria and check to see whether correct people matched. That's your most important thing. Eyeballing it and seeing whether it makes sense.

Enrollment data, which is prone to transcription errors and name changes, would make a good candidate for probabilistic linkage. When you're creating a patient index, think about the purpose of the file and what your ultimate analytic aim is. Then compare your results to previously reported benchmarks: Do they meet the smell taste?

Once you've created your MPI you can use this to link to a number of different sources like administrative health claims, electronic health records, vital statistics, and others. Both deterministic and probabilistic techniques can be used to link between data sources.

One thing I hope to quickly highlight right now is our linkage to the National Death Index the CDC maintains, which is a compilation of state death records. Researchers can apply to use the index, and if approved researchers send the NCHS a file with the identifying fields using the structure they specify. NCHS then matches the state research file with the national death results and returns a series of results files. Results in the NDI linkage will be complex, with multiple files and multiple linkages with a single person. The CDC provides guidance on how to interpret your results. This is a reference document you can have.

In our last few minutes I would like to go over selecting NDI data using a few of the examples I found in the literature. In one case, people used the NDI data to validate data that was in the Cancer Registry. What they recommended was that people get the most generous of the NDI files, which is the combined file, which lists many common, multiple matches for each record chosen. Then they use 15 criteria, as you can see here: Exact match, exact Social Security and sex match, exact match on Social Security only, 8-digit Social Security, 7, 6, 5.

As you can see here, they're kind of using probabilistic techniques although they're saying it's deterministic. They're not saying it's deterministic but they have used the date of birth within one year or something like that. To let you know, this information is in the Web Access Strategies for Using a National Death Index and Social Security Administration for Death Ascertainment in Large Occupational Cohort Mortality Studies and it's available with a link.

The results from this are that people get a lot of different files, different records. They get many different records meeting the criteria, then they review those, and as you can see for number 8, they only kept 21 out of 25. People went and reviewed things and saw if it's the same person, and they got a pretty good match rate from that.

This final example is a pretty interesting one from the state of Massachusetts where one of the legislative aims was Chapter 55, where they look at the effects of the opioid epidemic. This is pretty interesting because they took public health data, claims data, MassHealth (Medicaid), mental health data, prison data, and veterans' services and they aggregated it all and they linked them all. What they did in their linkage to understand the impact of the opioid linkage was to do only deterministic matches, but then look through the non-matches and try to link those together using these 10 different criteria. They're looser with the linkage criteria. Here's some examples of these referenced. Now I'll turn this over to Bruce.

Jessie: Thank you. Very helpful. Bruce Greenstein will speak from the perspective of his time in leadership positions in Louisiana and now from his role as CTO.

BG: I'm glad I got to be on this webinar. I just learned a lot as well. It reminded me of a lot of the work we did when I was in Louisiana. I'm going to divide my presentation into three parts. One is about HHS and

what we're doing now on these issues. I'm going to talk about my time in Louisiana and some complexities I'm sure everyone can relate to with regard to the structure of state government in helping human service departments. Then we'll discuss practical applications, not just creating a common client index or MPI but the real value that can be created when matching programs together, and the key is to key on those critical pieces of information to make sure you're talking about the same people in the program.

First, my perspectives come from being a CTO now and also when I was secretary in Louisiana. I did a couple tours at CMS as well, one from a regional office perspective and then one running a Medicaid managed care program in waivers and demonstrations. For some of my time I was on the NASHP board. I take all the work I've done and the brilliant people I've worked with and tried to put it in context to try to figure out the next steps we can do moving forward making the data we have for Medicaid and other health and human service agencies more valuable by connecting them and then having a business case that we execute on.

Let me start out with HHS. This is a beautiful view of probably the second ugliest building in Washington, D.C. followed ironically right after the HUD building, which is a little less attractive than ours. To remind people, HHS is really not this department that is perfectly unified but really a confederation of very large agencies with almost 80,000 FTEs and about as many contractors as well. You all know CMS, CDC, FDA and NIH, but there's also some other agencies that are not quite as big but play a very large role in state government organizations and create complexities themselves by carrying out different programs and having different rules around how data can be stored.

We are in the Office of the Chief Technology Officer. This position reports directly to the secretary, and therefore I get to work across the entire portfolio of programs within HHS. So we often have a role of bringing together multiple agencies to share data, and I'm going to use this as a practical application later in the discussion, for example between CMS and Medicaid data and CDC birth records. We do this for a variety of agencies. Last year we ran an opioid Code-a-Thon and brought together over 70 data sets from across the department, which offered us a glimpse into the complexities of data use agreements and how difficult it is to share data outside of each of the agencies, and then the very difficult task, if it can be done, to connect data sets using some kind of MPI across multiple data sets.

In our office here, very briefly, we have three business verticals. We have a pretty expansive portfolio. I'll draw eyes to the right-hand column, people in state government. We run an entrepreneur residence program and we often take people on IPAs from state government that want to spend one or two years in Washington, D.C. where they have permission from their home agency and state capitol and want to come to D.C. to learn and take on some large projects that give them a skill set that they bring back to their job afterwards. So I'm always in recruiting mode there as a former state official myself.

So we transition to the second piece around what our programs are like in Louisiana and how just the mere organization of our HHS-like agency, in this case really a health agency, had an impact on the way that we did our business. I was there in 2010-13. The population was 4.6 million people. Interestingly it's within 100,000 of the size of New Zealand today. It has a strong governor model, which means that the cabinet agencies operate under the governor normal but the budget and the way the authority is done is pretty strong relative to the Legislature compared to other states that many of the operational aspects require legislative approval. This sometimes is good to carry out executive-level tasks, but at the same time it also could cause friction with the Legislature, so working with them of course is best.

Within this department, it used to be called the Department of Health and Hospitals, now the Louisiana Department of Health. It includes Medicaid, public health, behavioral health, aging, developmental disabilities, but importantly it doesn't include all the human services. So our agency ran a very robust eligibility system as did the Children and Family Agency that took care of welfare, HeadStart and a variety of other programs.

We maintained, at the same time, these parallel eligibility systems that were in varying levels of preparation. Either a new system was being developed while running the old one or we were putting duct tape and baling wire on an old system and keeping it updated with the various programs that were being brought up or new populations being added along the way. That meant that we never had the perfect version of the data that we wanted to do either research or match across multiple agencies or do things for the purposes of efficiency only. All that, while in our state, and it's not different from most others, we were going through some significant budget pressures, which meant that we didn't have a whole lot of extra money to make big investments in our systems. But also during this time, CMS and HHS gave thumbs up for 90/10 matching on eligibility systems, which we certainly took advantage of, knowing that we needed an update.

At the high level in Louisiana, I worked with a number of states when I was in and out of government, and I have to say that the department in Louisiana had one of the strongest, most competent teams both on the Medicaid side and eligibility side. Not a ton of turnover, which meant we developed very strong skill sets. Aside from the technical skills, there also were great people skills, which is maybe just as important as certain interfaces in a computer system. That cultural attachment that allowed different agencies within the department to work together was also critical.

Our priorities also probably not unlike many other states. We continued to think about reducing provider burden, whether it means bringing up a new managed care program or new ways of doing prior authorization of different types of services, maybe a PDMP for opioid prescription. Every time we do something for efficiency or safety, it often means that there's somebody administrative or clinical out in the field that has to take an extra step to carry out these plans. Our goal always was to be very, very mindful of that, and often even though we were mindful of it, usually groups went to the Legislature and we would get hammered in committee hearings about the extra steps and times that it takes, and we weren't typically paying providers more to handle that. So it was always in our mind.

We wanted to make sure that as we rolled out these systems that we weren't asking it to be carried out with a lot of manual labor, a lot of new hires being required. That was something that we knew we weren't going to get, so we tried to make our systems be as automated as possible and create faster workflow for our people in the field.

We know that along the way programs change, whether it's Medicaid expansion or just new eligibility criteria or requirements for approvals, something in the system, and that also adds to the complexity of our eligibility systems. Then of course the state as we move forward there are always new laws and programs we were bringing up.

During our time we moved to managed care from fee for service in Medicaid on the behavioral health side we went from a carve-out to a carve-in. We ran new programs within public health where we tried to stop paying for healthcare services in our public health clinic with state-only money if people were enrolled in Medicaid. That seems like a pretty simple problem but it turned out we needed to match quite

a bit of data at the point of service in a delivery system that was generally just treating everyone for free that walked in the door. We knew we were leaving money on the table by not either charging or making sure that we were having Medicaid health plans pay for those services.

Then there were other programs in other departments in the state that we had to match people. You just heard a lot from Truven about methodologies used to ensure that either you had a high-integrity system, so we would work with our Department of Justice and our own Program Integrity Unit to be sure that we were not say paying two, five or 10 times for the same service, that somebody that was say dead we were not continuing to pay for capitation, which was fairly straightforward. Then there were also some providers that liked to bill when people were no longer in the state or alive. We actually were partnered with Truven during my time there and deployed many of those techniques you heard earlier.

Today the way we look at data at HHS is trying to move from a program to an enterprise data approach, but the same goes in a state government agency. Usually you have these eligibility files, which you try to make the single version of the truth to run the rest of your programs on. In Louisiana they call it a duolopoly. We had hours on the Medicaid side and there was also one on the Children and Family side. You want to know can you use that Medicaid ID outside of Medicaid? Can you use it to say payment integrity questions in public health or in other parts of the healthcare delivery system? Maybe it's in your psych unit hospitals or programs like behavioral health.

We talked earlier about an Enterprise Patient Index. Sometimes it's called a Common Client Index. It's been a bit of a Holy Grail for states for many, many years, being interested in having one unique identifier that all systems can run against. That logic is not too difficult but actually the system integration becomes very difficult. We still run very siloed systems and they don't tend to work easily together. Many states are working on this and some are actually doing this pretty well. I'm going to get to ways to try to address it.

In terms of how do you think through those problems, how do you make those issues make sense, I want to put your head into a way of thinking about it. First, the way to envision this is leading with the business case, why are you trying to do what you're doing? Why bother bringing data, connecting or matching it, and then creating an outcome? We just mentioned a couple of death data issues because you want to protect the integrity of your program. We just showed an example of Massachusetts opioid data mapping that's actually for many states, I would say Massachusetts and Indiana are very good examples to look at where they've made some investments in legislative action, in harmonizing state laws, but also within their state government to bring together multiple programs' data so they have a very accurate view, in this case of opioid overdoses and deaths, knowing precisely where they are and knowing some statistics.

An outcome of Massachusetts that I thought was interesting is they figured out that for people with a nonfatal overdose, one out of 10 will go on to die within the next year, which is awful, terrible. But it also means nine out of 10 are living. And if we can use some statistical algorithms who are that one out of 10 that is like to or does die, then we can undertake heroic interventions to try to save that person's life.

We have birth data and claims data in different parts of the agency. If you run vital records in Medicaid—and those can be brought together, which I'll talk about—and also within public health you have immunization data run in a registry. Then you have health plans, which are reporting quality. I want to spend a minute on these two examples on birth data and immunizations. One area that CMS and CDC are interested in is report carding out the outcome of births. Birth outcomes aren't completely captured in

Medicaid claims data. Medicaid would like to know how well they're doing. We're working on a project with CMS and CDC today to look for states that are matching their Medicaid claims data with their birth outcomes vital records data. We have seen so far some very good examples. Top of mind would be Ohio and New York, which are doing very fast matching, and the rich data fields from the birth record itself are then used to report on the quality of birth outcomes. Therefore, plans can be held accountable for quality metrics and states can be sure they're getting the most out of what they're paying Medicaid health plans for.

For immunizations, for example, most states around the country use immunization attainment as a quality indicator for Medicaid health plans. But yet they ask the health plans to self-report out how well they did based on claims data. We know there's an error rate associated with claims data and immunizations, and the much better data comes from state departments of public health, their immunization registries. With fairly straightforward matching—not super simple but fairly straightforward—reports can be created and then sent back to Medicaid that really give a very good idea based on how each health plan performs by ages, gender, parts of states. Those are insights that we never really use in general. Probably some states are doing a great job at it today. But those are insights that are available to us that we really could use moving forward and creating value.

The last thing is vital records and death data. A lot is made on how we obtain these data, how we pull out data. In general it's about how we fetch and exchange data that we think about. I'll remind people that I feel like historically we're on the cusp of something big. There's provisions around open APIs, and it requires those that house data that individuals should be able to get to that data with "no special effort." This means that we'll start to see a variety of APIs and the one that seems to get the most attention is the fire API standard. That's going to be a way that we start to push away the tribalism about keeping data locked up in systems, and then every time we want to go and get data out of a system we have to do these tribal acts. It's almost like having cups of tea and asking for special permission instead of creating the standard around the size of the plug that everybody can plug into, and then with the proper permission get data out.

We just talked about a couple business cases to do it, but as you can envision within your state, there are many more that create value, allows us to measure whether it's quality or efficiency, and allows us to get closer to assuring the highest level of integrity of our programs. I'd love to hear more what states are doing if you have good examples or particularly difficult challenges you're looking to address. Here's ways to get in touch with us. My email is: bruce.greenstein@hhs.gov. Thank you.

JP: Thank you. It's important to note that problems with eligibility can be addressed from a methodological perspective by analysts and programmers, but also leadership can make a really big difference, and there's a lot of work that can be done here.

Please submit questions via the chat box. Tracy will facilitate.

Tracy Yee (TY): We have some questions from our audience.

For Bruce: *Can you offer some insight into how to get public health and Medicaid working more closely together?*

BG: A great question. I just came back this afternoon from Atlanta working with CDC officials and digital bridge officials on lab reporting, as an example. It's not easy to do and there is a fair amount of tribalism

or insularity that is an actively passive way of not engaging. In other words, it takes leadership and championing. There's not a magic bullet that does it. As I mentioned, when you have people that break down the cultural barriers because they know one another, that tends to lead to better collaboration. I'm from Seattle and know let's say the woman that runs the Medicaid program is great and the woman that runs the immunization program is great. They have an easier time working together because they know one another.

In other places where maybe that relationship doesn't exist to start with or they're not in the same building or same agency, I always like to lead with that business use case. So a lot of it is where public health has the data that can show Medicaid an outcome of their program. That often would be appreciated highly. And to collaborate with one another to get to exactly what those measures look like or to get it, let's say, on a quarterly basis where Medicaid gives an eligibility file feed with a breakdown of health plan segmentation and then public health can turn it back around within a certain amount of time, with that sort then all of a sudden you're creating value. So I like to lead with need and that business case. I think that goes a long way to find a champion.

TY: We have a technical question so I'll direct it to Manju. *Can you identify a specific package or technique you would use for probabilistic matching?*

MG: I wish I could. I don't have a recommendation that I would use. I don't know if there's a single one size fits all. It really depends on the people who are going to do the work, what your research goal is, as Bruce said the business need for it, and how technical you are. I do know that a lot of vendors sell products which make it easier to match. However, if you have a very SAS-based shop, I know that there's a lot of SAS Macros assumptions which mimic probabilistic techniques, like Soundex and some SAS Macro packages that were developed by CDC called Link+ or University of Washington called Blink King, which uses Proc SQL. But if you don't want to move away from SAS, you can locate library shareware support for R.

Also I just wanted to mention that probabilistic matching is a statistical technique. People are starting to move a lot more into machine learning and having the computer figure out which things go together and then reviewing whether 'hi' means the same as 'hello' or things like that. So I think we're going to see a lot more push into that area.

TY: For Bruce: *Is there a place for health plans to participate in data sharing or exchange between Medicaid and public health?*

BG: Very good question. There's not a place place, like some sort of clubhouse we all get together and discuss this. But there ought to be. There are a couple organizations that have done a pretty good job of assembling it on the back end for research purposes: the Healthcare Cost Containment Group here in Washington, D.C. Optum Labs has massive amounts of mainly health plan data but some Medicaid. States really are the center of attention and focus. So some states have been doing this kind of work for a long time and are probably good for mentoring other states that have a keen interest in infomatics around it or on these matching processes to get to some very positive outcomes.

Like I said, in some of our background work looking at birth outcomes, I know that Ohio and New York have done some extensive work. There are a number of other states that are making a lot of progress. I'd like to see this as the norm rather than the exception over time, and my office, given that we work with

each of the HHS agencies, is definitely willing to be a convener to bring together parts from states and from the federal government to help.

JP: To summarize some brief takeaways from today's webinar:

- An awareness of administrative issues with Medicaid IDs can improve the success of an analytic project or at least demonstrate that an analyst or programmer has the ability to avoid a lot of possible traps with an appropriate approach to common challenges.

- We provided an overview of probabilistic versus deterministic matching and some of the considerations to take into account when choosing between these two approaches. Probabilistic matching may be the better choice depending on your analytic project.

- Constructing a positive return on investment for enhancing records systems can help support evidence-based policy decisions. Encouraging data linkages at the leadership level can create more robust state data environments and allow for more enhanced program monitoring.

Thank you for joining us. Please complete the feedback survey. For more information or for questions directed at the IAP Data Analytics team, contact us at MedicaidIAP@cms.hhs.gov with Data Analytics in the subject line.