Medicaid Innovation Acceleration Program (IAP)
Using Geospatial Statistics to Analyze Medicaid Data
August 31, 2017

[Logistics]

Jessie Parker, CMS: I'm the government task lead overseeing the IAP Data Analytic technical support work. I'm also an analyst in the Center for Medicaid and CHIP Services. We have two speakers today who bring a lot of experience in the area of geospatial analysis:

Marty Jolly, Team Lead Programmer Analyst, Government Health and Human Services, Truven Health Analytics. Marty will walk us through an example of a hotspotting project on GIS cases in North Carolina.

Aaron Truchil, Director of Analytics & Informatics, Camden Coalition. Aaron will be preventing on some of the exciting work that Camden Coalition has done on healthcare focus through geospatial analysis.

Today's agenda is:

- Overview of Medicaid Innovation Accelerator Program (Jessie Parker)
- Introduction to Approaches for Geospatial Analysis (within the context of our example geospatial project)
- Example of a Geospatial Project (Aaron will walk through other examples of how geospatial analysis can be used in a healthcare setting, including recent work done by the Camden Coalition)
- Camden Coalition's Use of Hotspotting
- Questions and Answers

This webinar is present by the Medicaid IAP. IAP is a cross center collaboration at CMS intended to support safe payment and delivery system reform efforts. We are in the data analytic functional area. One of the approaches we are using to increase data analytic capacity within state Medicaid agencies is webinars such as this one on data-related topics and challenges we think might be relevant across a wide array of states.

In this interactive webinar, states will learn about:

- Map types and usage
- Key questions to ask when evaluating a map
- Geostatistics
- Hotspot maps
- Key components of effective geospatial projects

The goal is that by the end of this webinar you will understand the basics to do spatial analysis and will start thinking about how you might be able to apply this methodology in your state Medicaid program.

Our first Speaker is Marty Jolly, who will walk us through an example of a healthcare geospatial analysis project to lay the groundwork for understanding this methodology, including an overview on hotspotting. He will also discuss what return on investment you might expect from implementing this type of analysis.

Marty Jolly: My role is lead programmer analyst in Truven Health Analytic's federal consulting, now part of IBM Watson Health. My work revolves around collecting healthcare data, forming that data into data

set reports and geo visualizations for federal and state government. In my career I've done two major geospatial projects – a bank ATM placement and retail store placement, along with mapping for Medicaid SUD and 1115 waivers. For clarification, let me speak on the terms GIS versus geospatial, both of which I will use in the webinar.

The term GIS, or geographic information system, started being used in the early '80s when computer programs enabled the creation and analysis of digital maps. When using the terms GIS, I'm speaking about these programs and related systems. The term geospatialist started being popularized in the early 2000s. Technology such as satellites, GPS, smartphones and geolocated transactions have pushed people to find a broader term like geospatial. I will use geospatial when speaking about modern technology and more recently deployed analytic approaches.

Geospatial analysis is a broad and deep subject. To help me stay on track I've structured this presentation around an example of a geospatial project. We'll start at the beginning of the project and quickly cover each step. As we move through the project I will highlight key steps. Let's start with first issues in Medicaid geospatial projects. Let's look at the things we might put on a Medicaid map and questions that may be asked. An effective geospatial project will start with a question. It doesn't have to be a great question. Geospatial analysis is an iterative process. You start with one question, produce a map, and voila you now have multiple better questions.

Once you have a question in mind, get your things on the map: Regions, providers, beneficiaries, expenditures, services. After you've proved to yourself that that map makes sense, you ask that first question. Maybe the question is what is the spatial distribution of my withdrawal treatment facilities in the state? Once the map is produced to answer that question a new question may be asked, like what is the spatial distribution of my population by income level? This process could continue ad infinitum and is at the heart of geospatial analysis.

We should also think about the team required for effective geospatial analysis and key skills to emphasize. We need direction, project management skills, deep subject area expertise, comprehensive data governance, computer GIS skills, and statistical knowledge. All of my projects have had a team size of four – a stakeholder who gave direction, a project manager keeping the project on track, a PhD statistician, and a GIS analyst programmer. Subject area expertise was shared between all team members and data governance was mainly handled by the GIS analyst. On larger projects the team size would increase to provide depth in subject area expertise and data governance.

There are many factors that make up an effective project. One factor I'd like to emphasize is data governance. Data governance is a loaded term at the enterprise level, with the goal of ensuring that data can be trusted. Here I will only touch on a specific piece, maintaining data integrity. Geospatial projects add another layer of data, the map data. It is this additional layer that makes data governance especially important. As you iterate your questions and maps and time shortens to deliverables, maintaining the integrity of data is crucial. For example, if I started a geospatial project managing the data manually in Excel, and the project gained traction, I wouldn't want to switch my data management to a programmatic environment like SAS, where data manipulations and joins could be codified to ensure auditing and repeatability. You do not want to be in the situation of Excel hell, where at the end of the project you're manually reorganizing the data to a new map satisfying last moment epiphanies.

Let's briefly discuss ROI for geospatial projects. Return on investment is an important subject, cost-benefit. My time today only allows me to bring attention to a well laid-out book, The Business Benefits of GIS: An ROI Approach (David Maguire, Victoria Kouyoumijan, Ross Smith, ESRI Press – 2008). This book provides a framework in which the key questions can be answered, answered in ways where the value to

the stakeholder is identified, benefits are measurable, and standardized techniques for estimated value can be established. We can imagine ways to apply GIS technology to Medicaid challenges, such as possibly using geospatial analysis to be more efficient with interventions. But I hesitate in making specific examples. Conversation with your Medicaid stakeholders should be held identifying their most pressing concerns and then thinking how geospatial analysis can be applied to reduce cost and improve performance.

Now our example of a geospatial project. The example project we will see is based upon an analysis of North Carolina SIDS – sudden infant death syndrome – for the years 1979-84. This analysis was described in the paper cited below, written by the geospatial scholars Arthur Getis and J.K. Ord. ("The Analysis of Spatial Association by Use of Distance Statistics" by Arthur Getis and J. K. Ord, Geographical Analysis, Vol. 24, Issue 3, pages 199-201) The scholars Getis and Ord, among others, such as Luke Anselin, have developed key geostatistical techniques used in GIS for things like hotspot maps, which we will cover.

Let's look at the North Carolina SIDS project in detail. Getis and Ord were given the North Carolina births and SIDS deaths by county and started out with the assumption that there was no spatial association in the data. We will see how their geostatistical approach provides an objective identification of hotspots with confidence levels. As mentioned, a geospatial project will start with a question, then there's the search for data. We have our question: Is there a spatial association between counties and SIDS events? Now let's take a closer look at the North Carolina SIDS data provided.

I found the North Carolina SIDS data used in this study on the University of Chicago Center for Spatial Data Science website, cited below. We have birth SIDS counts along with the SIDS to birth ratio. It is this ratio that is used in the analysis. You can see the ratio-defining columns F and G. For this presentation, we will only be looking at the data for 1979. The underlying statistics require a minimum of 30 analysis items, and there must be a significant variance in the item values. Also any known spatial associations must be taken into account. For the North Carolina SIDS data there are 100 counties with significant variance of the SIDS to birth ratio values, and there are no known spatial associations in the data, such as any North Carolina regional viral outbreaks during that time.

Let's take a sidebar and dive into some key points around data in general. There's a wealth of geospatial data generated by government at the federal, state and county levels. Census.gov is a great place to start exploring this data. There are two basic types of geospatial data sets. I call them map data and data data. Map data are all the information necessary to put something on the map. You will see this referred to as the Topographically Integrated Geographic Encoding and Referencing/Line (TIGER/Line) map data. TIGER files are commonly stored in a format called shape files, which are easily loaded into GIS programs.

Data data describes those things on the map. This is like our North Carolina SIDS data. On the systems.gov site, there is data from American Factfinder and the American Community Survey sites, which also can be joined to a map via geographic identifiers. Some files you may download just have the map data, some just the data data, and some will have the map and data prepackaged together like the link here of TIGER/Line with selected demographic and economic data. Additionally, you have data data, state-managed data with geographic identifiers. You can join this data to a map for analysis.

Back to our project. We have our data data. Now let's get our map data. I'll delve into the 1990 North Carolina County boundary map, TIGER shape file from census.gov. We loaded it into a GIS program, and we see the North Carolina County reference map labeled with county names. A reference map just shows the things being studied on the map, in this case North Carolina County. Now is a good time to take a

sidebar and dive into map details, specifically map projections. Every choice made in creating a map is important. Making bad choices can lead to confusing, even misleading maps. A first choice is what projection to use for the map. I hope to share that picking the correct projection for Medicaid-related maps is critical but doesn't have to be difficult.

Maps are created using an inside out perspective. On the projection shown here, a globe with latitude and longitude lines is wrapped with a sheet, and a light is projected from within through the globe onto the cylindrical sheet. Once laid flat, we have the map. But you will notice in this projected map, the further from the equator, the more distortion of shape and area. We see here the area of Greenland appears to be similar to that of Africa, which is definitely not the case. Taking an outside in approach, using a basketball postcard stamp analogy, we can imagine trying to form the postcard to the surface of the basketball. It cannot be done without folding or tearing and distorting the postcard. But you can apply the stamp to the ball with insignificant distortion. The point being, if you're centered over a small enough region, distortion is minimized. In this analogy the postcard is the national level and the stamp the state level.

Let's explore more of this focusing on Medicaid maps. This slide compares two different projections along with an unprojected map in green. The projections were centered just west of Topeka, Kansas, the circled dot. We see that the map of Kansas for both projections are very similar, but for other states like Montana there are significant differences in aspects of shape, area, and direction. For our type of maps, it's the distortion of area that is the most concern. If we were to center over Billings, Montana, the projection differences for Montana would be significant, but then the other states would be distorted. The point here is for our North Carolina SIDS analysis, if we center the map over North Carolina, the map projection differences become insignificant. To make choosing a projection simpler, each state has identified projections that fit our type of analysis. I go to the state GIS portal and find the projection they're using for the type of map I'm creating. If in doubt I will contact the state GIS department. If you're interested in a deeper understanding of projections, this link [on slide] is helpful and has a table comparing projections and this usage in great detail

Back to our North Carolina SIDS project. Recapping, we gathered our North Carolina SIDS data data and the North Carolina County map data. Now we need to join the data to the map. Using a GIS program, we see here the North Carolina SIDS ratios were joined to the North Carolina County reference map. We are making progress, but it's still hard to see any spatial association of the county's SIDS ratios. The next step is to produce a map attempting to get an idea of spatial association. There are many ways to accomplish this. We will look at a chloropleth map, commonly known as a thematic map.

Thematic maps are a good way to start seeing any spatial patterns in the data, but they can be misleading. There are many options when creating a thematic map. Here we see one option called Equal Count. Note in the legend in bracket we see the count in counties in each color shading group. The equal count option ensures that there is the same number of counties in each color group, and in this case the darkest red group represents groups where the SIDS ratio is 2.81 or higher. Thematic maps also allow the viewer to get a subjective idea of spatial patterns. An excellent primer on how maps are misleading can be found in Mark Monmonier's book "How to Lie with Maps." A more statistically rigorous analysis is needed. For now let's see how selecting different thematic options appear on a map.

Here the equal interval option was selected. This option ensures the range of SIDS ratio values are the same for each color group. We can start to see how depending on how the option selected, the thematic map may drive different conclusions depending upon the person viewing the map's different interpretation. Now let's look at the thematic option Getis-Ord used in their paper. The Getis-Ord

thematic map appears similar to the equal interval option but is different in that they chose ranges following the 1, 2, 3, 4 color groups. Getis and Ord were not satisfied with the subjective nature of thematic maps so their work addressed that issue. Let's look at their hotspot map, which applies geostatistical rigor.

Using the hotspot map feature in a GIS program, I was able to create a hotspot map very similar to the map in the Getis-Ord paper. We see in the legend where hot and cold spots are identified with associated confident levels. The thematic maps use a descriptive approach based upon color shading groups, where the hotspot map uses the statistical approach based upon probability theory. Practically speaking, I view this as eliminating the noise, the noise being statistically insignificant counties. From an analytic perspective, hotspot maps identify statistically significant areas where further study may be applied, in the SIDS case by epidemiologists and social scientists. From an economic perspective, questions like if I had one dollar to spend where would that dollar be best spent can be ignored. The Getis-Ord approach is to get a more local indication of hotspots in relationship to the entire study area.

To get a deeper sense of how it works, let's dive under the hotspot hood. I created this simple map for the purpose of explaining how hotspot maps work. What you're looking at is the square study area of 100 features. Let's call this the square state with mostly square counties. Each feature or county is surrounded by a neighborhood, in this case adjoining counties. Each feature has an associated attribute value. I randomly assigned attribute values between 1 and 100, then changed a few values to create hot and cold spots. The process simply put is for each county the average attribute value of the neighborhood is calculated and compared to the average of the study area, which gives us a ratio. Also the probability that the neighborhood's distribution is not random is calculated. In statistical terms, the ratio is the Z score and the probability is the P value. Once this is done for each county, we can produce the hotspot map showing counties that have statistically significant hot and cold values. This is how we get the more local indication of spatial association following the geospatial axiom everything is related to everything else but near things are more related than distant things.

I would quickly like to mention the tools used in creating this presentation. For the map side of things I used the open source GIS program QGIS. For the data side Excel was used. Researching GIS statistical algorithms was done using Internet resources like the web portals, stack exchange, and GIS Lounge, along with Google Scholar searches.

Finally, I hope I've given you a sense that geospatial analysis is iterative. Once data is mapped, more questions emerge, having visualized the data spatially. All maps are misleading in some sense. The options chosen for thematic maps can affect our perception. Thematics are good starting points but are very subjective with differing interpretation. Hotspot maps are generated with statistical rigor, providing an objective analysis, highlighting areas where focused study may help evaluate underlying causes. Data governance in geospatial projects is especially important with the addition of the map data and its own set of handling requirements. Thank you very much.

Jessie Parker: Thank you for that great overview of geospatial analysis. That was a lot of complex information in a short timespan so we'll open the floor to questions before the next speaker. Please submit questions via the chatbox and Tracy Yee will facilitate the Q&A session.

Tracy Yee: We have a few questions. The first one is a two-parter focused on data governance: How does data governance relate specifically to geospatial analysis, and related to that are there areas that I need to be concerned about compared to other types of analysis and reporting?

Marty Jolly: I struggled over using the words data governance because it's such a loaded term. I'm speaking from my experience. The important thing I can bring out from that is because you have another

set of data, the map data, and the speed at which geospatial projects will iterate, being able to join that data properly and being able to go back and audit what you've done and be reliable, the processes you put in place for managing the integrity of that data is critical. I'm not sure I answered the question.

Tracy Yee: Another question regarding the SIDS study in North Carolina: While the scholars used the equation of looking at all births over SIDS deaths, they didn't look at all deaths over SIDS deaths. Why didn't they take total deaths into account for this analysis?

Marty Jolly: I read their paper and did not see anything speaking to that but I know that this subject was of great concern and was looked at. CDC did a huge effort on that part. I don't know if I can answer that question. I don't believe there was an exact determination of the cause for SIDS deaths, but the education efforts for how the parents put their child to sleep, not prone but lying on their back or side, that contributed to the reduction in SIDS deaths.

Tracy Yee: A question regarding your toolkit: Besides ESRI, are there other GIS software programs that can populate hotspot analyses?

Marty Jolly: Yes, absolutely. ESRI is the enterprise level GIS software, but the Oakland source tools such as QGIS and GeoDa…in fact, Luke Anselin, who I mentioned in the webinar, developed with his group the GeoDa software, so there's a lot of work coming out on the open source that I would definitely recommend people taking advantage of if they don't have the resources at the enterprise level.

Tracy Yee: One question is you mentioned that hotspot maps eliminate the noise. Are you suggesting that we not use thematic maps when approaching geospatial analysis?

Marty Jolly: No, absolutely not. Thematic maps are the first place people start, but as we saw, there's two things: There's a lot of room for interpretation and then just a lot of noise. To the hotspot maps take that noise out by you don't see the statistically insignificant counties in our case. But using thematics and hotspot maps, I think you use them in conjunction.

Tracy Yee: An audience member would like to know doesn't a county at the edge of the state boundary become difficult to find an accurate hotspot for it?

Marty Jolly: Absolutely, there are a lot of considerations in there. The edge cases are very important but practically speaking, in reality, with jurisdictions and data collection methods, all those things have to be taken into account and you end up at the end of the day making the best analysis that you can, factoring in the fact that on the edges of your study area there could be significant factors. But we live in a real world and we live with limited data and resources. That's something we have to work with.

Jessie Parker: Building on Marty's example of applying geospatial statistics to identify county hotspots in the SIDS outbreak, we will next hear from Aaron Truchil from the Camden Coalition. Aaron will be discussing some of the innovative work the Camden Coalition has done using hotspotting to identify and analyze healthcare costs and utilization.

Aaron Truchil: Marty gave a really great foundation for understanding geospatial work and highlighted some interesting and relevant scholarly work in this space. The work we're doing here at the Camden Coalition is a nice complement in that it shows some interesting applications of using geospatial analysis for on-the-ground community-based work. You saw a lot focused at the county level and we're going to drill down even further and really get into pretty granular level data.

Before I dive into that a little context for those not familiar with the Camden Coalition. We are a 15-year-old organization. We're not a Medicaid agency. We're a nonprofit. We're a membership organization. New Jersey took a relatively interesting approach to the Medicaid accountable care space, where instead

of creating agreements between hospitals and managed care organizations (MCOs), we created a model called a geographic accountable care organization (ACO) through the ACO demonstration project. What that means is that all the hospitals within a defined geography need to be part of the ACO as well as a large percentage of the primary care practices. There's a variety of other considerations.

The Camden Coalition itself is a community ACO for the city of Camden, New Jersey. We're located in southern New Jersey, a population of roughly 80,000 people. The majority of the population is within the Medicaid space. So we're providing a variety of clinical services, predominantly community-based care management and care coordination. We do a lot of work with our partner hospitals and primary care practices around quality improvement work, and all that is supported by a variety of different policy and data and research activities. Geospatial is one piece of what we're using data for and I'm really excited to talk more in depth about that today.

The Camden Coalition's data work started by taking the three health systems that are physically located in our city of Camden, New Jersey, three hospitals that have long histories of competing and not collaborating with one another as well as they could be, and bringing them together under one common data sharing agreement. We were able to pass data sharing agreements with each hospital to enable us access to their billing data, that data being individual-level claims for emergency department and inpatient encounters. Taking that data, we were able to take the identifiers from all those different data sets and link them so that say, Aaron Truchil at Cooper could be connected with Aaron Truchil at Lords, and we could start to get a more global picture of how Aaron's hospital utilization looked, incorporating all the different, traditionally disconnected data sets.

We took that data set. It was citywide, all-payer. It wasn't just Medicaid, Medicare or commercial; it was every single encounter at the hospital. We linked that up and created a really powerful data set that allowed us to see a lot of interesting trends about the city that we're working in. We found that for a city of roughly 80,000 people, we spend roughly $130 million each year in hospital-based care. That's predominantly Medicaid and Medicare dollars. Of the 80,000 people living in the city, roughly half, a pretty high percentage, are going to the hospital each year for an ED visit or inpatient visit. Even more importantly, we saw the extent to which there was overlap, so individuals that were visiting multiple health systems within the same year or potentially over a few years. That was really some of the impetus to create that community ACO framework I mentioned earlier.

A lot of our work that has become more publicized is focusing in on the outlier population. So if you took everybody's cost of their hospital-related care and looked at the distribution as you see here, this is something that's become much more discussed as of late. But this is the sort of core essence of our work in that we see extremely high concentrations of costs in small subsets of the population, and we look at this as a really profound opportunity to understand what is going on in these individuals' lives and start to try and make some improvements that we think are going to have a really disproportionate impact because there's so much concentration of costs. So a lot of our work has been focused in on the 1 percent of the population that's driving roughly 30 percent of our spend.

What that looks like on an individual level basis, this is one case that we pulled out of the data. This is based on a real case but it's been anonymized and shifted around, but this is a 5-year timeline based on the hospital claims data and being able to see each of those ticks as a distinct encounter with the health system. So if it's red it's an ED visit, if it's blue it's an inpatient encounter, and you can start to see the sequential spacing of them. There's some diagnosis grouping so you can get a sense of the progression over time and the complexity. All that is to say is these are really expensive, profound cases where there's lots of touchpoints and lots of money being spent but we're not seeing the underlying causes of the trend

being reversed. So a lot of our work has been really trying to unpackage what drives it and how can we fix it, and geographic analysis plays one interesting role in that.

What does the 1 percent look like in aggregate? We're talking multiple chronic conditions, on average more than five. In the Camden population that's averaging around 57 years old, in a given year a little more than 4 ED visits, 5 inpatient encounters, someone who spends 54 days total per year in the jail, and wracks up an average of roughly $75,000 in just hospital-based receipts. This doesn't include other aspects of care as well. You start to get a sense of what that individual looks like. That's really been the foundation of hotspotting.

We spent time in Marty's presentation talking about hotspotting and people have come to use hotspotting in a lot of different ways. As it applies to geospatial analyses it has a very particular definition, but as the Camden Coalition has operationalized it, a lot of it stems from the work that was happening in New York in the police department and criminal justice space over the '90s. You can read about that in The New Yorker article that came out that Atul Gawandewrote a handful of years ago but this is just to give you a quick sense of how we operationalize hotspotting. So hotspotting is the strategic use of data to target evidence-based services to complex patients with high utilization, so this is really about identifying patients that are getting a mismatch between what their needs are and what services are available. Data is a really essential piece of that.

Now that I've given a high-level conceptual view of what our work looks like, let's dive into some of the geospatial techniques that we've applied. As I mentioned earlier, our foundational work started with hospital claims data. We didn't have Medicaid claims and eligibility data; we had the data available to us, but in many ways there's a lot of similarities. When we think about geospatial work, you can take your table of claims data – deidentified or anonymized claims data – and there's a lot of opportunities to do interesting analyses around this. If this is something new to you but you have a baseline familiarity with data and graphing, essentially what a map is is you're looking at a scattered plot. On one axis latitude, on another axis longitude, and now if you can start to think about a third dimension, it's easy to have some degree of intensity or some other way of being able to show an additional data point. So when you start to look at a map, you can start to see that now we have neighborhoods on the map, and we can have concentrations or hotspots or clusters starting to jump off, and there's lots of different ways you can define that, but this is really the foundation of what mapping looks like at its more raw level.

So tapping into the claims data that we had, of course we had individual level data, so not only names, dates of birth, social security numbers that allowed us to link individuals across the system, but we also had the addresses that individuals were reporting. So address, zip code, city, state – that allowed us to do a technique called geocoding, which converts those addresses into latitudes and longitudes, which then allows you to place that address on a map. Through that process we're able to take the claims data and convert it into a geospatial map. So this is just one example of that. This is showing all the hospital dates and spending in a given period. You see about five years here but at the census block level, so this is pretty granular. We're talking a few city states connected together. You start to see that in a place like Camden, New Jersey, there's lots of concentration at both the individual address level as well as the block level. So you can start to see from that data certain buildings or blocks that are responsible, just as we showed that cost curve at the individual level, certain geographies that are responsible for a disproportionate share of our spending. Here in Camden we see that 6 percent of our city blocks account for roughly 18 percent of our patients and almost 40 percent of our hospital spending.

Again these are opportunities to start to understand what's going on in those buildings and blocks that's driving it. Is it just a concentration of people? Is it some kind of geographical—some place-based strategy

that needs to be deployed so it starts to bend that cost curve? So a map is really a starting point to get the conversation going, to understand what the community looks like a little bit better and start to get the idea of potential strategies. It doesn't point you particularly down the path of what you need to do, but it can start to really help stimulate your brainstorming, your engagement of partners. So what we've done over the years is start to build relationships with a lot of those buildings and communities that jumped out in those early maps.

What you see here is one of those buildings. It's a senior living facility and over a 5-year period that we're studying we saw roughly $12 million in hospital spending. That's obviously a lot of money. We could start to think about what kinds of strategies might make sense in that building and would those strategies result in reduced hospitalization so we could start to get at the question of ROI. We could start to tap into the diagnoses codes in the claims data and start to get a building level health profile and understand what are the top diagnoses that are driving different types of visits. Then even going a step further, we can start to drill in and understand the population within that building.

We were able to sign a data sharing agreement with the social service agency in the building that allowed us access to their individual level data, which we were then able to link into the hospital data. What that showed was that there were different populations even within that hotspot building. So you could see the majority of the population was rarely if ever going to the hospital over the handful of years of this study. There were different profiles of individuals, some who had pretty frequent ED visits, two, three or even more per average a year, and you see the reds, a very high number of ED visits and not too many admissions. Then you see in the green another small subset of individuals who had pretty high degrees of inpatient stays. They weren't going to the ED as frequently but were getting admitted and readmitted to the hospital frequently.

We could start to think about what kinds of strategies and approaches made sense. Was it intensive care management and care coordination for that high inpatient utilizer cohort? For some of the individuals with lower ED visits, maybe predominantly ambulatory care center visits that could have been handled by the primary care setting, is it about engagement in trying to strength connections with primary care providers? There's lots of different ways you can take this. The geographic analyses laid a really strong foundation. One for us to prioritize around was where are the buildings or neighborhoods that we should be focusing our efforts on, but simultaneously how do we engage those different stakeholders—say the social services or the residents of that building? In this case we were going around the building and talking to residents or an extensive focus group in a listening session where we were presenting maps and data to them and saying do you think you're getting good quality care for $12 million over multiple years of hospital spending? It's a really good technique for trying to build relationships and rapport with a lot of different partners. That was really important for our work.

That being said, trying to identify high-utilizing populations is just one application for how you use geospatial analyses. As I said, that claims data we were starting with, there's a whole variety of different interesting applications that could be done. Over the years we've been approached by various partners that have had different interests and supported them in a number of different ways. I just wanted to highlight a couple of them and then talk briefly about some work we've done on the Medicaid side.

Here you see a map of falls. Falls are obviously very costly hospital-related incidents, so there's lots of efforts to try and reduce the number of falls. It's an injury that lends itself to a place-based intervention, so starting to think through where are the places we're seeing the highest concentrations of falls. This is based on where individuals are recording as their residence so it doesn't necessarily mean where the fall is happening, but you can start to imagine that the majority of falls are going to happen in an individual's

home. So through mapping the data we can actually start to see where do we see really high concentrations of individuals that are actually showing up at the hospitals with falls, then we can start in the same way I mentioned earlier to build relationships with the buildings and the residents and start to think about what strategies might make sense. So you can see here there were a handful of buildings. Those buildings tended to have this is what we expected, a higher concentration of senior citizens, which would lend itself to that. But the data played a powerful role in confirming that, helping quantify it, make the case for why there's potentially an ROI, so it was a really interesting application.

Similarly we were interested in different pregnancy rates and quantifying the costs of pregnancies. We were able to take the diagnoses data and be able to map that for a separate project. This is really just to say that there's lots of interesting things that lend themselves – you heard Marty talked about SIDS, so there was a tremendous opportunity – you're only limited to the extent in which you want to be creative and how you think about these things.

Briefly, as our work around hotspotting has grown and we've partnered with lots of other communities across the country, we've gotten opportunities to work with a full variety of different partners and apply some of these techniques and approaches in other communities. So one of those early partnerships was with Mainecare, the Medicaid agency in Maine, and Maine Quality Counts. We entered into a data sharing agreement with them similarly where we got Medicaid claims data for a 3-county range. We picked a few counties up in Maine. I have a link here to the report. You can see we applied the same level of analyses trying to understand who were the inpatient high utilizers and the ED high utilizers or the common trends that we can pull from them and how do we start to map them and see if there's any kinds of geographic trends. What you see here is just zoomed down at the county level.

Really what was interesting about the Maine analysis was we've done the majority of our work at least up to that point was in Camden, a very dense, northeast city, a very concentrated area. Now we're applying it to a much larger area that included rural stretches and some cities. Obviously you've got Portland here. But a lot of other rural and suburban communities as well. You can start to see basically these density maps that there were intense concentrations and that concentrations of costs levels still holds true. It might look a little different in rural settings but by focusing in on certain subsets of the population you can start to see that you can have a disproportionate impact. So of course Portland being one of the most populous areas jumped out, so we can drill in. These are just a small subset. We produced a whole variety of different maps that went down to the billing level but this is what we were publicly releasing, and you can start to see it drilling into Portland, various communities that were intense pockets of high spending and high utilizer concentrations, and that really helped guide some of the initial work on their end as they started to think about if they were interested in building out community-based care management programs and how would they deploy them, what neighborhoods and communities made the most sense?

So we can take this data first and foremost presented on a map but underlying all that data are lots of rich tables, so if you're pulling census data or eligibility data that gives you a sense of what is the denominator. You can start to get a sense of which towns have the highest concentrations and get a sense of what are your proportions. So there's lot of really interesting potential analyses that can come out of this.

A few takeaways before we open it to questions:

- From our viewpoint, GIS is just one of many tools in your analytic toolbox. You're going to be doing analyses in Excel or our statistical package that's comfortable for you. GIS is just one step in that journey of trying to understand what the story is with the data. So whether your angle is to

produce a map or build some tables, you should always think about how you can incorporate GIS in.

- GIS and geospatial analysis isn't just limited to just high utilizers and I showed a few examples. I'm really interested if there's particular questions or applications of interest to the group to discuss them in more detail. Really, I think it's about trying to understand what's relevant to your existing work so where do you have existing efforts that might benefit from geospatial questions or a geospatial lens.
- Lastly, Marty has emphasized what his team structures have looked like historically. The Coalition offers an interesting complement to that in that we really started from a much more grassroots level. We didn't have a tremendous amount of costs. We were trying to use low-cost or open source software. We didn't have PhD-level staff working on this. So this is something you can start small with and start to build your capacity, whether that means training your existing data analyst in GIS, and there's a whole variety of online curriculum to do that. As it relates to being able to show some baseline descriptive data and put some maps together, those are all things that you could start to build that capacity over time and then to do the level of sophistication that needs to be used to pick up some of the statistical significance that Marty mentioned, yes, you might need to engage outside researchers, but there's a lot of low-hanging fruit along that way. So don't think about that you either need to build the gold standard shop that's a lot of different people. There's a lot of opportunities to start much more simply.

Jessie Parker: Thanks. That was a good and interesting presentation. It's been useful to see multiple ways in which hotspotting and GIS analysis can be used in the healthcare setting. We're short on time but any last questions.

Tracy Yee: We have a couple questions from audience members. The first one is if there are any implications or challenges to using address data for low-income populations?

Aaron Truchil: Yes, that's something we often need to ask ourselves as we embark on this journey: What are the biases and assumptions or some of the constraints that face us? If you've been working in the Medicaid space I'm sure you're pretty familiar to the degree to which there's address instability, whether individuals moving around or housing instability. So your ability to make inferences is only as good as your underlying data. That being said, you have to ask yourself that, and there's lots of different ways to assess the quality of your data. There's opportunities to explore that and have more comfort in that. We've actually taken an interesting approach in that having addresses and understanding that instability is one additional marker that we might want to look at. Can we see in our data individuals bouncing around across different addresses? That's one more marker that might be driving excessive hospital utilization.

We also bring in additional data sets, things like shelters, and we can cross reference if individuals are reporting shelter-based addresses, which might give us a proxy for homelessness. So you have to do the constraint. It's also an opportunity to think about some other interesting questions as well.

Tracy Yee: A final question is you mentioned that geospatial analysis is not limited to high utilizers. Can you briefly discuss other applications of this methodology?

Aaron Truchil: Yes. I presented on a few quick tidbits looking at where you might see fall hotspots. It's still looking at where you're seeing spatial concentrations but instead of looking at, say, high utilizers as I showed earlier, multiple hospitalizations, multiple chronic conditions, these were just incidences of falls

and where you might want to deploy a place-based intervention. There's a lot of interesting scholarly work and some of it's been done and there's still opportunities to do it. You might want to look at where you see pockets. You could look at what's called network analysis, an individual's distance between their home and place of care and how that impacts their outcomes. So really the sky's the limit. You have to think about if we have diagnosis data, which diagnoses might we want to understand from a geospatial perspective? Or if there's a quality metric you're looking at, which one might lend itself to a geographic analysis? So hotspotting can be applied to so many different questions that it's really just what's meaningful to your community as you're working through your priorities.

Jessie Parker: Some key takeaways are summarized here and we're at time so I won't spend it here. Benefits of geospatial analysis include:

- Relevancy to Medicaid questions on capacity, access, utilization, and expenditures
- Useful for identifying areas of interest for further analysis
- Relatively low cost when executed by data analysts with open source software

But I think a key takeaway is that not only can this help save your agency money by aiding with targeted interventions, but the analysis itself can be relatively low-cost and executed by analysts with open source software. In a low-resource environment as many of us operate in this is important.

Thanks to the audience and presenters.

We'd like everyone to complete the post webinar survey. Your feedback is extremely helpful and I actually read them all myself. For more information on the IAP program or questions, reach us at MedicaidIAP@cms.hhs.gov. Slides and a session recording will be posted on our Medicaid IAP Data Analytics website.

[end of tape]