Cyberseminar Transcript

Date: March 20, 2019

Series: Patient Aligned Care Teams (PACT) Demonstration Labs

Session: A Population Health Informatics Approach for Measuring Obesity Prevalence among VHA Patients

Presenter: Hadi Kharrazi, MHI, MD, PhD

*This is an unedited transcript of this session. As such, it may contain omissions or errors due to sound quality or misinterpretation. For clarification or verification of any points in the transcript, please refer to the audio version posted at* [http://www.hsrd.research.va.gov/cyberseminars/catalog-archive.cfm](file:///C:\Users\VHASLCMyersK\AppData\Local\Microsoft\Windows\Temporary%20Internet%20Files\Content.IE5\73RXHSNF\l)

Moderator: I would like to introduce our speaker. Joining us today we are lucky to have Dr. Hadi Kharrazi and he is an assistant professor at John Hopkins School of Public Health in the Department of Health Policy and Management and has a joint appointment in the John Hopkins School of Medicine Division of Health Sciences Informatics. He’s also the research director at John Hopkins Center for Population Health IT. So with that, I would like to turn it over to you now.

[pause 00:26-00:34]

Moderator: Looks great, thank you.

Dr. Hadi Kharrazi: Okay, thank you for the introduction. So as [unintelligible 00:40] mentioned, I am Johns Hopkins School of Public Health and what I’m presenting today is a collaboration with had with the Office of Analytics and Business Intelligence at the VA at the time that was the term used. Now it’s CSDE and also the PACT team, the Patient Alignment Care Team on measuring obesity prevalence among VHA patients. This was sort of a topic of interest in the broader domain of population health and how we can use informatics resources to better understand what is happening to our VA patients.

So here is our first poll. I’d really like to know who the audience are and what is your primary role at the VA.

Moderator: Thank you. So for our attendees, the poll is up on your slide, I mean on your screen, right now so go ahead and click the response most closely related. Student, training or fellow; clinician; researcher, analyst or data/IT scientist; administrator, manager or policy maker or other. And it looks like about 80% of our attendees have responded. And that’s great so I’m going to go ahead and close these out and share the results. And you were able to select more than one options so our percentages will reflect that. So it looks like 6% clicked student, trainee or fellow. Seventeen percent clinician, 51% research, analyst or data/IT scientist, 20% administrator, manager or policy maker and 17% selected other. So those of you that selected other, please note that in the feedback survey at the end of the presentation we will have a longer list of job titles so you might find your exact one to click there. And we’re back on your slides now.

Dr. Hadi Kharrazi: Great, thank you. So before I jump into the actual project, I wanted just to briefly talk about what is population health and how we are using informatics or IT to empower population health wide interventions. Now knowing that most of you are researchers or administrators have seen this before. So this is triple aim. In addition to lowering cost through improvements in our health system and also improving the care for our individuals, there is a third aim here and that is how to improve the health for our population. And that is where our focus is at our center here at Johns Hopkins. And we were approached by the Office of Analytics and Business Intelligence as they were developing a platform on how to collect the data on population level, how to manage and evaluate the different subpopulations of the VA patients and what that means in the larger context of healthcare delivery at the VA. And in the midst of this of course we focused on obesity which I will talk more about it today.

So our Center for Population Health Information Technology at Hopkins, CPHIT, we pronounce it CPHIT, tries to improve the health of populations by advancing cutting edge health IT across all sectors. So healthcare providers is one of the sectors. We also look at payers. We also look at other source of data including social determinants of health and so on. The typical outcomes that we focus on are healthcare utilization outcomes like cost, hospitalization, ER admissions and others. And we usually developed a lot of models to predict these outcomes and our predictors are typical clinical predictors like demographics, diagnoses, medication, but also other sources of data including social determinants and geographical information. Now you can see the typical data sources we use are insurance claims, EHRs, health information exchanges, discharges and so on. And usually we look at large-scale populations which actually VA has that sort of a population, and usually across multiple years to see the trend. There’s the link to our website here as well which is jhsph.edu/cphit.

Now typically what we do in our science is we develop a lot of different platforms where we bring in all of the relevant databases, we try to link them together. That is sort of what you see on the slide on the top left side of the slide and after we have done this, we find issues. We have to fix data quality, problems we find, which typically uses 90% of our time. And then we try to do our modeling, extracting knowledge. We do the data mining. We define certain patterns that eventually we hope to be able to share with others and then they can use it. They can give us feedback, they can find their populations, we get new data and then the cycle goes on and on and on. So this is sort of a learning health systems cycle on a population side and how we do it. Now, in this project we are basically showing you how we got to maybe the extraction part before even I talk about the modeling. I will briefly talk about the modeling at the end, but we had such great results in the first phase of it just by looking at the prevalence of obesity, that was worth of presenting to the VA community.

Now usually when we create these models, we calculate risk scores for patients and then depending on what risk score our patient has for a certain outcome like hospitalization or something else, they would be sliced into different subpopulations. We all know the so called high disease burden patients are a fraction of patients that may need more direct involvement and management, like case management and care coordination. Things like that. But then as you go down the list of patients or subpopulations of lower risk, usually their numbers are bigger and they often need a lighter intervention as you can see here, like needs assessment and quality improvement for the majority of healthier users and non-users. So this is something that has been done in our center for many years. The results of all of these modelings have been culminated into an actual tool that is widely used in academia called the John Hopkins ACG tool, or advanced care grouping with a very special focus of course on the outpatient setting for disease management and risk scoring of the patients. And a number of our VA projects, we have also used that just to let you know that this is sort of the ultimate goal of population health management.

Here you can see we feed in demographics, medical service files like diagnostics and procedures. And pharmacy files, medications to be able to calculate the risks you can see all of the input files on the left. Then there’s a lot of modeling happening and on the right side it actually creates all of the risk scores that could be used for operational purposes.

Now in order to go beyond the [unintelligible 09:14] world where they have all of the data in one place and it’s often cleaner than what we have on the provider side. That becomes a bit tricky. So as you can see here, if you look at the center of this image, there’s where the patient and physician interact with each other and that’s always a fraction of time and each of these entities actually interact with their own circle. So you can see physicians on the left side they interact with the [unintelligible 09:46] practice team. They’re bigger network of their healthcare system like if it’s an integrated delivery system, it’s an accountable care organization and so on. And on the right side you can see the patient interaction with their family and caregivers and their broader communities. And if you look at this, sort of the big view of both sides, the patients and the providers, there are many data sources that could be used to better understand risk and outcomes for different patients or populations. EHR, or the electronic health record is of course one growing area of interest. Here you can see a circle actually showing it mostly on the left side, but there are a lot of other circles here like on the left side you can see the health information exchange, insurance claims, a lot of different national data sets that are available for different health networks and so on. But on the right side, you can see there is a personal health record and then there is all of these gadgets, mobile health apps and even social media, geographical information on where the patient lives and so on. So we have gradually in our center started doing all of the risk stratification based on the EHR data. The EHR data have its own challenges. It had a big of a leakage and so on. But that is how we started collaborating with the Office of Analytics and Business Intelligence at the VA, because they were in the midst of solidifying all of their population health platforms and understanding what type of data is needed in order to scale it. And within that, as an example, we picked up obesity to test to make sure that this platform would work.

Now here is sort of a high-level view on our collaborations that we have had with the VA. We did develop a conceptual population health analytic framework which was very academically oriented. We had to do systematic reviews of literature, find out what is needed, what are the different data sources out there, what the VA has access to and hypothetically how each of these data sources would help in predicting certain risk scores. Then we had to test that analytic framework in terms of its technical feasibility and of course that goes back a lot into the underlying systems that the VA has, such as the Corporate Data Warehouse or CDW. And we picked up obesity as a sample population because it’s across the entire population. The sample size is very big. And then we started actually looking at that and trying to come up with models to predict utilization in different segments of the overweight or obese population. And in the midst of this of course we started exploring the prevalence of obesity on a geo-temporal basis and we found some interesting trends. We are not epidemiologists, but we did work with them at the beginning of this study and we found certain trends that was both very informative but also helpful for us in terms of creating the models. In this presentation again I will only touch on the modeling at the end of the presentation, but that’s not the focus. The focus is to give you the findings and the prevalence. And of course we worked a lot with the PACT team, or evaluation teams on finding out what are different subpopulations of the Veterans who worked with the PACT team have received different utilization services or not. So that was also something that we brought back from this population health analytic framework.

So now focusing more the so called obesity prevalence project.

I have another poll question to ask you. I appreciate if you can open that.

Moderator: Thank you. So for our attendees, as you can see, we do have another poll question up on your screen. And we’d like to get an idea, what do you think about VHA patients’ obesity rate compared to the general population. And again, you can select all that apply. VHA population has a higher rate of obesity, VA population has a different trend of obesity over time, VHA has a different geo-distribution of obesity or VA population has a different set of factors associated, varied factors associated with obesity. Go ahead and take your time to select those that apply and it looks like about two-thirds of our audience have responded so far so we’ll give people a few more seconds to get those answers in. Okay. I’m going to go ahead and close this poll out and share those results.

So, sorry, 58% of our respondents selected option one, higher rate of obesity. Thirty-five percent selected option two, different trend of obesity over time. Thirty-three percent selected different geo-distribution of obesity and fifty-six percent selected the final option which is there’s a different set of varied factors associated with obesity. So thank you to those respondents and we are back on your slides.

Dr. Hadi Kharrazi: Okay great. Actually these were our questions getting into this project because we didn’t know and what we knew was basically from some publications that I’ll get to right away, but we needed to explore it ourselves.

Now just going back quickly to the population health analytic framework that we developed, we looked at these larger population health management framework and we came up with, as you can see it’s a very busy diagram, but basically there are two pieces in this. The left sort of boxes that are noted with a light green line around them shows that there are a lot of factors that will play a role in the health of a population and of course we can’t look at the genomic data [unintelligible 16:32-16:33] was not in that area, but we can always look at clinical factors or social and non-clinical factors like environments or lifestyle or social data. And that’s why we not only use the CDW, but also geographical information to find the trends in the obese population. But then also in the middle of the slide you can see there is a box, population health analytics [unintelligible 16:58] sort of denoted by a red line around it. And that’s where the analytics come in and then depending on what is the outcome or what are the operational management end points, then there are different analytics tasks that could be applied to it including things like risk stratification, coming up with scores and so on. And of course on the right side of the slide you can see all of those sort of outcomes on a population level that could be measured or effected.

Now, going back to what we thought about those questions that you just answered. We looked at some literature and at that time, this was the most recent one that we found. There’s also another publication that came after this, but at that time we had this. You can see that this publication actually did show a higher rate of obesity among our Veteran population, mainly the overweight class one and two were more predominant there. Unfortunately most of the publication didn’t have much information or reliable information on the female Veterans mainly because of the lower sample size that we have there. But based on what we saw we had to start from scratch. And the first thing again we are [unintelligible 18:33] develop risk models, we develop ways to find out what happens in the future to different segments of the population, so we did a quick prevalence check.

So we wanted just to see what our data is representative of what we are have seen in prior literature. So this is showing the percentage of underweight, normal, overweight, obese class one and two and three categories based on actually one day of the VA data. So this one day in 2013 had actually close to 30,000 visits with a valid BMI recorded on those days, on that day. So even one day, one slice of the VA data actually gave us numbers that are very close to what the data that was published earlier in literature.

Now this is also a very quick plot that we did in less than a week just on those 30,000 points just to see what is the geographical distribution of patients who visited on the VA facilities in that day. And as you can see, they are all over and you can basically see the US map here. Of course this is all of the continuous attached states together and the colors basically shows whether they’re underweight, normal, overweight or obese. So this gave us the confidence that the data is there and we can look at nor only the prevalence rates, but also the geographical distribution and potentially the temporal trends as well because that was only one day of data.

Now I don’t want to go into the details on how we cleaned up the data, but I can just tell you this. It took us almost six to eight months just to clean up weight and height data. Weight and height in the earlier years in the VistA EHR used to be apparently a free text area so there was a lot of typos issues, a lot of other things that we saw in terms of data quality. So we had to write an extensive set of rules to be able to clean the data. And we were dealing with, I don’t want to say big data in a computer sciency aspect, but we had a close to a hundred million weight and height records across all years that we looked across all of the VA patients that we had on record. So it was hard, [unintelligible 21:23] was definitely not a way that you could do this manually. This diagram basically shows how we [unintelligible 21:30] an actual weight of a person could have been modified until it gets into our modeling techniques. If you need to know more about how we cleaned up the data, I’m more than happy to provide you with additional information later.

Now we again wanted to do a quick test. We looked at the 2015 data because the study was actually conducted mid-2016, so this was the most recent complete year of data we had. We looked at the numbers, I don’t want to go through the numbers on this chart, you can look at it later, but it did actually follow the major trend lines of the earlier papers on obesity. So it gave us also sort of a quality check for us, making sure that what you’re looking at is reliable and comparable to other data.

Now in terms of our methods I don’t want to go on the how we did the modeling to predict how the BMI will change in different geographies and so on, but this is the method that we used to just come up with the prevalence rates. Of course we had to do a lot of data quality checks on the completeness of data, accuracy, timeliness, provenance and other things. Then we had to do a lot of adjustments for BMI. We know BMI does increase by age, so when we want to show a map of BMI you don’t want to just show where the elderly live versus the younger Veterans, so you need to adjust for age or adjust for certain neighborhood effects. So, we had to do multiple different adjustments depending on what we are showing on screen and you can see the list of methods here including two different way of linear model regressions because BMI is a continuous variable. We also did multi-level modeling because every neighborhood might have a different linear model associated with it. But also we looked at epidemiological direct and indirect adjustment methods. And then for the GIS you will see some slides with hot spotting results. We used the Poisson and Bernoulli methods to do that. Now I will also show in some of my slides, the results for the female Veterans but there is a possibility of skewedness there. I cautioned you on reading, not too much reading into those results because some of the neighborhoods don’t have enough sample and also some of the different categories that I will show it to you. So the male population is the more reliable data that we will show you. And a lot of these results are unpublished. If they are published I will make sure that VA will post them on somewhere that you can see it.

Now the results. So the results are the majority of the presentation. A lot of it are maps. I will just go through them with a faster pace because it’s basically, very much it’s visual. Now, the first thing was, you can see here on the left side are the males. On the right side are the females. Again the female population, there is always the sample size issue. We tried our best but sometimes it’s unreliable. I don’t want to go into that. This is more of a population level and how this data can help us to do better risk stratification. Now you can see on the males, this is age stratified. Basically you know every ten years you can see, we have from 18 up to 80+, the different age categories. And the associated line with it. The actual symbols used in each of the maps actually shows the order of these age categories when you reach to the right side of each of these diagrams. So if you look at the left diagram, the male population, you can see on the left side is year 2000 and then it’s 2005, ten and 15. So as you move to the right side of the males diagram it shows how the BMI is being increased. But, at the end of the day in 2015, the order of the those lines that you see matches with how the symbols are put on the map. So you can see the age category for 40 to 50 years old actually tends to have the highest BMI compared to the other categories on the male side. On the female side, again with that caveat of unreliability of our data, that is that’s sort of what the order is [unintelligible 26:34] different when you reach 2015. Regardless of that order between male and females or within the male population you can see across all age categories BMI is being increased. Now you might ask how much is it being increased? You can see the lowest BMI in year 2000 for the male population is around 28 and they’re sort of crossing 29, 29 and a half as you go to the right side and some of them actually cross that 30 threshold when you get in 2015.

Now I showed the same slide set here, but now stratified based on race. And you can see for example on the left side the male population, the Asian race actually has the lowest BMI overall across all years, but still the BMI is being increased there as well. Now, you might ask that this might be the effect of age. This is actually age adjusted. So it’s not like you know, people have aged by five years and they gained weight here. This is all age adjusted. And you can see the other races also listed on, shown on this diagram. The same thing for the females. An interesting part was that on the male part you can see Native Americans have the highest rate among all the races, while the African American Veterans tend to be actually lower compared to the Natives and the Whites and the most racial categories. On the females side that actually follows the CDC tend of African American actually Veterans having the highest rate overall.

Now you may ask about underlying population we looked at and everything else, all of those results are available and again these are unpublished results. Fairly soon you will see all of those results with details about the population that we have looked at. But overall this is the entire population of Veterans available in the CDW, the Corporate Data Warehouse between 2000 and 2015. The exclusion criteria has been minimal to things like pregnancies or certain surgeries or certain procedures that would definitely effect weight in an artificial manner. So those things, if you have any questions, I will be able to address at the end of the presentation.

Here is ethnicity. You know, Hispanic, non-Hispanic. You can see the trends on the male side. The fluctuation on the female side is for example, for multi, ethnicity is mainly because of the low sample size so it’s when you see a lot of fluctuations across the years, that tends to show a low sample size that we have.

Now here’s the self-reported income. I stress this is self-reported income. This is not an income based on different ways to figure out the SES, or the socioeconomic status of a Veteran. We know the self-reported incomes are sometimes skewed but you can see the results for males and females and the differences there as well. To our surprise, the lowest income category of 0-25k actually had the lowest BMI trend. Again there are issues with correctness of self-reported income or accuracy of it and we are actually looking at other ways to evaluate that right now.

Now marital status we can see again males on the left side, females on the right side. And you can see for males, the married group actually had the highest BMI overall and for females that was the widowers. Again I stress the fact that no matter what side you’re looking at, male or female, or what category you can see all categories regardless of the age adjustment are being increased.

Now here’s their service. You can see we have Persian Gulf. We have World War II, Post-Korea, Vietnam, Korea and so on. These are categories used in CDW, not always the best categorizations for the service location. You can see also again the same trend happening. And I sort of give a warning again on the female diagram there’s a low sample size issue that some of the lines actually I haven’t provided [unintelligible 31:35] p-value and everything else. But some of the lines and some of the points are not reliable. We can’t actually make a judgement on that.

This is the last sort of trend line analysis here on the branch that they served. And you can see Navy and Marine actually are higher on the male side and the Coast Guard is actually sort of the lowest. And to our surprise actually Air Force was in the middle.

Now the map. This is where we actually used the geographical information of where a Veterans lives to find out what is the average BMI for that so called neighborhood for Veterans living there. And of course there’s a lot of data cleaning happening here. A lot of statistically adjustments to make sure that it’s a good representation of that neighborhood. You can see some of, this is a county-level map. I cannot show you the maps for the female Veterans because of a lot of the data quality and low sample size issues we have. Again, if you reach out to me I might be able to give you those specific areas that we think we have good representation. But for the male Veterans we had enough sample actually to show the entire country expect from some of these countries here that are greyed out because we didn’t have enough samples there. You can see that the range actually goes from 27, that’s the greenest color you see, to around 30ish which is yellow and then 32 which is red. Remember 27 and higher, based on CDC classification means that you’re already overweight. So green doesn’t mean here that they’re really healthy, but it’s sort of they’re overweight, but then crossing into yellow and especially going into red, that’s bad. Now this is the male population and it’s adjusted based on a multi-level modeling sort of adjustment for age. So as we go through the different maps, so this is 2000. I’ll show you 2001, two, three, up to 2015. Again, I stress the fact that this population is not aging because everything you see is age adjusted. So I’ll start going through these slides. So this is 2000.

You can see in 2001 some neighborhoods actually are turning more into yellow and sort of orange. And I’ll stop speaking here. I’ll let you to really see what is happening to the maps. So this is 2002, 2003, four, five, six, 2007, 2008, 2009, 2010, 2011, 2012. You can see that some regions of the country are really contrasting with the other regions. And then we have 13, 14 and 15.

So it is sort of disappointing to see that overall in these 15 years the BMI is being increased regardless of the age of our Veteran population. But to our surprise, the main region with the highest increase in the BMI is not the South, very much like the CDC map. It is actually the upper Midwest. And there are areas that the increase has been less than the other areas and that’s sort of the mountain areas around Colorado and New Mexico and so on.

So overall this is what we see on a county level. The upper left side is 2000 and as you go row by row from left to right, at the bottom right you see the end result. And that was alarming for us of course and this is very important that for certain maybe areas there needs to be more attention given to the obesity interventions, like the MOVE! Program and [unintelligible 36:12] others that VA has rolled out over the years compared to the other parts.

So of course you can look at this on a state level, but a state level doesn’t give you the granularity of pinpointing what is happening where. This is for the male population. I can, although again there is a sample size issue, I can give you a rough estimate on what the map looks like for the female population on a state level. I cannot give you the county maps, but you can see actually it is not very dissimilar than the male population, but this actually follows the CDC map of sort of the Southern states having a higher obesity rate. We actually expected this for the male population as well but we were surprised that for the male population was different.

Now it’s all about the localities, so although we see these regions and as a policymaker you might think about different allocations of different resources to different regions, but it’s all about inner city, suburb and the rural areas. You can see here, I zoomed in the DC Baltimore area, so what you see, that enclosure up there and the image on the up right side of this map is the Baltimore area and DC is of course down to the left, you can see that the inner city actually, this is for males. The inner city actually looks more healthier and again the green means they’re already overweight, but they’re not like in any of the classes of obesity. And as you move out to the suburbs and the rural areas, it gets worse. And we saw the same trend almost in a lot of other urban areas.

So this is, for example, Philadelphia. You can see inner city has a more healthier overweight population compared to the obese population living in the more suburb or the rural areas. I’ll just give you some other urban centers or MSA, the metropolitan statistical areas.

Like this is New York. You can see the city and the surrounding boroughs.

This is Atlanta. There’s more of a contrast here because the shift to the rural areas and that suburbs happens faster in some of our cities compared to the coastal ones.

This is Chicago and some other areas close to Chicago. Just to give you an estimation of the fact that everything is about the locality here.

And this is Houston. So you can see again inner city versus the rest of the area.

Now you might say this is prevalence, what about hot spotting? And we did that as well. So in hot spotting we are measuring whether there is a certain legion of, or certain neighborhoods, have a higher probability of having obesity cases in it compared to controls. And it actually gives us sort of a p-value and confidence interval and sort of tells us what are the probability of being obese in one neighborhood compared to other ones. So on a national level you can see it does follow a bit the prevalence as well, but then you see some hot spots that you couldn’t see in the prevalence map. So for example, in some of our Northern states there are hotspots that are pretty high in terms of the probability of being obese by living there. So it’s almost two times higher in some of these hot spots. Now again if you drilled down into certain areas such as the VISNs here, you know VISNs could be informative to those regional VA headquarters of understanding their population.

So here is, for example, the mid-Atlantic and basically if you look at the map on the right side. The left side shows obese versus non-obese, but the right side shows the difference. And you can see there are certain areas that are healthier than others. And that is very important to know if you have more of a VISN-level policy.

This is another VISN for example. The Heartland Network. That’s Illinois, Kansas and Missouri. You can see how the spread is. There are here some very specific points that I will talk about it, why there are certain specific points that are hotter than other areas.

Here is the South Central. Note that not all scales across all of the VISNs that I showed are the same. This is one to five, so that point that is five is really high [unintelligible 41:42] the likelihood of being obese just by being there.

And here for example is almost 30. It’s very, very high compared to the rest of the state. This is actually the Texas VISN.

And here is the Rocky Mountain one.

You might ask where are those hotspots and we actually spent some time on trying to geographically make sense of where those hotspots are. We created a lot of interactive visualization techniques to be able to look into it. And at the end of the day a lot of those hotspots that I showed that are very high on the scale, like five, six, seven up to sometimes in rare cases like 20, 25, the likelihood of being obese compared to the random distribution of the VA patients, they mainly overlap with Indian reservations. And that actually confirmed all of our racial categorization or stratifications of BMI as well, so that would be informative not only to the VHA but maybe IHS and other agencies that can address these issues in specialized subpopulations.

And we also created visual analytics just to get a sense of if, in almost real time, if we want to look at these distributions what does it mean? And can we update these? Remember I started the entire presentation with the population analytic framework, so we wanted to see [unintelligible 43:27] technically we can always update these things [unintelligible 40:30] almost on a nightly basis or not. So we only prototype these things and we hope that one day these become operational. There are a lot of different databases listed on this slide that we’re used for visualizations and some of the real-time analytics.

This is the only slide I want to show you on the modelling. You can see we experimented for example with the modelings in the mid-Atlantic VISN to see what effects the obesity rate. And you can see we have variables like age, race, income, marriage, service years and so on. And [unintelligible 44:11] you can see that certain things like a Gagne score might be effective or highly associated with obesity rates, but when you start adding other variables like social determinants of health like income, marriage status and also the geography like urban, road, low food access and so on, the effects might be a bit different. The p-value shows whether it’s significant, the effect or not. So the ones that are less than 0.001 or 0.05, you can look at those rows. And then the odds ratios are also listed there, whether something actually increases the likelihood of being obese versus not. So if you want to know more about the modelling projects we did, I’m more than happy to provide that later.

Now just a couple slides for the discussion and then we will open up any questions. Oh, we have another poll here. So I appreciate it if you can open the poll.

Moderator: Thank you. So for our attendees, as you can see we do have the third poll question up. In you opinion, what was the most time-consuming task in this population level analysis research? Please go ahead and just select one option this time. The options are, cleaning weight and height data, attaching the geodata to underlying populations, developing models to predict the obesity trajectory or understanding the factors associated with obesity. Pardon me, it looks like just over half of our audience has replied thus far, but the answers are still coming in so I’m going to give people a few more seconds to get those in.

[pause 45:56-46:03].

Okay, it looks like we’ve head just around a 70% response rate so at this point I’m going to go ahead and close the poll out and share those results. So 76% of our respondents selected cleaning weight and height data. Five percent, attaching the geodata to underlying populations. Sixteen percent, developing models to predict the obesity trajectory and three percent, understanding the factors associated with obesity. So we’ll close that out and we are back on your slides.

Dr. Hadi Kharrazi: Thank you. Actually the audience is right. We spent a lot of time cleaning something that we thought, well we cleaned to start with. You know, something [unintelligible 46:49] weight and height. And it was very time consuming on our end because there is no turnkey solutions that would clean everything for you. That’s more of a need for more solutions in that area although computational powers needed for any of these modelling techniques was a bit of an impediment, but overall the most appropriate answer is actually cleaning weight and height.

Now, a couple discussion points. When we do population health analytics we always need to pay attention to what data source we are bringing in and whether that data source has any value or not. I know at the VA for example we have a lot of new claims coming in because of the Choice Act and whether that data is useful [unintelligible 47:39] in addition to the EHR data we will have, that’s a big question. Eventually DoD data will come in under the new unified EHR platform that will happen in the near future so they’re all questions about that and how that would effect anything on a population health level. Data quality I talked enough about it so please pay attention to that. And also the denominator reflection. So I didn’t talk much about how we came up with our denominator, I just gave a couple criteria on how we excluded certain small, narrow populations, but sometimes as you start trimming your population to get to your study population, you may actually lose your signal instead of trying to get the noise.

Now challenges of using the EHR data, such as CDW is that VistA or any other commercial EHR that may replace it, they’re all designed for transactional applications and data quality is always a challenge. But on the other side, the good thing is they’re a reliable, not necessarily a valid source of data for a population health research. For example, you don’t need to run [unintelligible 48:58-48:59] wait and hide. I mean wait and hide is being stored anyways. And they’re real time data. This is very important on population health management. Instead of planning on a five-year basis or an annual basis, the planning could actually come down to monthly or weekly basis. And it also saves a lot of effort in data collection and other things necessary at the VA.

Here’s a couple of VHA potential collaboration. One is we are interested in bringing in the social determinants of health on the either personal level or geodrive level in order to find out whether there are any disparities to control for certain mediators, moderators. Identify certain patient subgroups aka patient phenotyping in order to enhance either population level analytics or even clinical research. And then, as I said, additional data sources, especially the DoD data or the claims data coming in that is of high interest to us. Measuring the effect of data quality on the research outcomes is of interest and also risk stratification of very specific subpopulations of the VA patients such as mental health or especially our female Veterans that often have sample size issues.

So just wanted to acknowledge the teams. On the VHA side, at the time again it was the Office of Analytics and Business Intelligence, Dr. Fihn and Dr. Box. And also the VA PACT team. This is actually the long list of PIs on the VA side. More than happy to share the entire list with you later, but they were all around the country, around like 12 people. And also you can see the folks from our center, the Center for Population Health IT.

So last question before we open it up for your questions.

Moderator: Thank you. So for our attendees, as you can see on your screen you do have the final poll question. And again, you can feel free to select all that apply. The answer options are, developing technological infrastructure considering electronic health record migration, assessing different value of data sources, geodata, claims data etc. Identifying and then targeting disparities, for example with mental health or females. Funding future studies to exemplify population level analysis. And looks like just about a third of our respondents, I’m sorry a third of our audience members have responded so we’ll give people some more time.

[pause 51:48-51:54]

All right, we’re looking at just about half of our attendees have replied. So go ahead and get your responses in and wrap this up. Okay, I’m going to go ahead close this out and share those results. Sixty-six percent of respondents selected developing technological infrastructure, 34% assessing different value of data sources, 59% identifying and targeting disparities, 38% funding future studies to exemplify population level analysis. And with that I’m going to go ahead and put it back on your slides and we can move on.

Dr. Hadi Kharrazi: Thank you. This is very helpful to know your answers on what you think should be the next step. We are actually engaged with the new Office of Analytics and Business Intelligence, has a new acronym, on actually all levels here and we hope to be able to take it to the next stage. We know that the VA in the in the midst of all of this EHR migration and a lot of things happening on the background, but the need has never been that high especially now that the Choice and MISSION Acts are passed and we are dealing with a more, I don’t want to call it fragmented, but a more diverse type of data coming into the VA. So this was the presentation. I’m more than happy to send you any additional information as needed. You can see my email listed here and also the link to our center. Thank you.

Moderator: Thank you very much. So for our attendees that joined us after the top of the hour, you are able to submit questions and comments at this time. Please use the GoToWebinar control panel located on the right-hand side of your screen. Down towards the bottom you’ll see a question section. Just click the arrow next to the word question. That will expand the dialogue box and you can then submit your question or comment there. We did have several come in during the presentation. People are looking for a copy of today’s slides. You do have a link to those in the reminder email you received about three hours ago. There’s a live hyperlink in the email from HSRD Cyberseminar or you can write in to the questions box and I can send you that link.

The first is a comment that came in. Given that we serve Veterans, factors like military exposures, theater of operations, military occupations, should also be included in the framework. Do you agree?

Dr. Hadi Kharrazi: I fully agree. It’s all about, so one problem we have in population level analysis is data quality and [unintelligible 54:34] completeness of data. So when the completeness of data falls below a certain level we can’t make a good judgement of whether the trend is reliable enough or not. I just want to also put a caveat out for the [unintelligible 54:50-54:51] on the call or the methodologists on the call, that we are talking about millions of patients. So a lot of times p-values are not that reliable. P-values were designed for small samples of clinical trials like 50, 60, a hundred, maybe a thousand patients. And when you’re in millions always the p-value is significant. We are talking about effect sizes and whether that effect size has a meaningful operational outcome. So I definitely agree. I wish if we had access to those data as well.

Moderator: Thank you. Looks like lots of good questions coming in. Is there a publication we can cite with some of these data? And if not, when one might be available?

Dr. Hadi Kharrazi: So [inaudible 55:40] publication is not available right now. We had to go through certain clearances which we actually are expected to receive pretty soon. [Unintelligible 55:51] as we receive it, we will submit it. You probably will see the, considering the reviews and things like that, probably end of summer is when the publication will come out.

Moderator: Thank you. And as a follow up to that, would you be able to share any VISN-level data for grant applications at this time?

Dr. Hadi Kharrazi: Sort of the quick answer is yes and no. So yes on my part as sort of as a Hopkins faculty member, we are more than happy to do that for you. The little no is I need to get confirmation from the PI from the VA side that they’re willing to share that data. But please send me an email. More than happy to sort of connect, hopefully getting the permission so we can share the data with you.

Moderator: Thank you. The next question. In hot spot analysis, did you use age-adjusted rates as well?

Dr. Hadi Kharrazi: Yes, we did. That is correct. If you don’t do that we are basically showing where are the hot spots of elderly among our VA population. Also, a little note. The mean age of our Veterans is already close actually to 62 or three on the male side. So I just needed to put that caveat out, that overall the BMI is higher than the general population because the mean age is higher anyways. But yes, it was age adjusted.

Moderator: Thank you. The next question. Impressive geomaps. Did I miss Puerto Rico or was it included?

Dr. Hadi Kharrazi: So the prevalence maps for Puerto Rico and all other, Alaska and other territories and islands are available. There are a couple issues with sample sizes in some of the remote islands of course. We can share that, but we just wanted to show the continuously attached states. Now on the geomapping, there are issues with islands in general and sometimes the results are not that reliable. I need to double check that on the hot spotting, but we do have prevalence maps for Puerto Rico.

Moderator: Thank you. What was the GIS software used?

Dr. Hadi Kharrazi: So it was a mix of different things. So the entire analysis is done in R, which is an open source statistical package. All of the rendering of the maps were also done in R. The analysis of the geo spotting is R plus an external software which is funded by NIH. I can share the link and everything else. It’s a freely available open source software that you can download and actually works in tandem with R to do the calculations. For some of the internal visualization purposes, not necessarily the ones that I showed, we also used ArcGIS. The problem was that because we had so many points, tools like ArcGIS could not do some of the calculations, hence we used R or other scalable sort of software that can do this.

Moderator: Thank you. Has there been any data collected or researched regarding impact of having patients take shoes and socks off when having height and weight taken?

Dr. Hadi Kharrazi: So there was no data on that. So that is definitely a data quality issue. So you know, you have one patient actually being weighted three times a day and every time it’s a couple pounds off. So we actually had to look at a lot of different thinks like what are the minimum and maximum physiologically possible weight based on Guinness Records? What are, sort of, we call that hard limits. Then there are some soft limits. What are the outliers? And then this is not only on the population level, we also need to look at the patient level on a temporal basis. What are the minimum maximum of a change of weight during a given period of time. So it goes on and on and on, on how to clean it to get to the bottom of it. Of course, we developed almost 20 some rules to clean it, but frankly on a population level on such big maps, the first couple rules will clean it enough. Of course, there are a lot of typos, issues with that and so on. Rounding issues. You find out that, for example one center puts everything in a rounded numbers for some reason but another center didn’t round them. But rounding could be also on tens and fives as well. So there are a lot of things there, but we would never know the truth because we never know what was the condition of the weight scaling and we never know whether the actual scale was accurate or not. So we just rely on what data we have and on a population level on a national map like this, those little differences won’t make a [unintelligible 1:01:35] of a huge change.

Moderator: Thank you. We are at the top of the hour, but we just have two pending questions. Are you able to stay on and answer those for the recording, or should I send them to you offline?

Dr. Hadi Kharrazi: No, I’m available.

Moderator: Wonderful, thank you. If any attendees have to drop off at the top of the hour when you close our of the meeting please take just a second to fill out the feedback survey. It’s just a few questions, but we look closely at your responses and it helps us to improve individual presentations as well as the program as a whole.

The next question. What basic trainings does the analytisist [sic] need to have in order to conduct this type of analysis?

Dr. Hadi Kharrazi: So there’s sort of two pieces in the presentation. I talked about the risk stratification and so on. That’s a very much of a population risk management and modelling technique and I had only one slide at the end sort of showing one of the preliminary models we had. That’s very biostat-oriented, machine learning oriented. But the majority of the content that I showed you requires a lot of knowledge on epidemiology, demography, how to adjust and some modelling for the geography. So it needs a bit a geographical information sort of a knowledge of GIS systems and how to do geoanalysis or geo-temporal analysis I should say. So it’s a mix of things and we actually had a good number of people with different skills sets that everybody contributed to each piece of it. The short answer is probably one person cannot do it. You need a team. But I tried to give you some of the skill sets needed.

Moderator: Thank you. We do have a comment that came in. [Inaudible 1:03:39-1:03:40] also been performed by other HSR&D researchers, NCP, VSSC, MHSP and the Center for Health Equity and the National Center for Health Statistics. Has the NCP vetted some of the numbers presented here? It might be worth it to review recent population level studies compared to the findings with what has been found in the VA.

Dr. Hadi Kharrazi: No, it has not been vetted and that is one of the reason that we are waiting. But not NCP. So I’d if you, whoever provided the comment, I’d appreciate if you can send an email to me, the email is on the screen. I’d love to know who could actually help us on vetting some of this information that we have generated. But we did work with a number of people at CDC that run the BRFSS and other survey based you know, and some of it it’s also actual measurement of weight and height and how they come up with the national maps for the general population so to vet the data and we try to get country level and maybe census-track level information to see what our census maps are all about. You know, inner city versus suburb and rural [unintelligible 1:05:07] what we saw here. So we had some discussion there, but again if you have any connections to people who have sort of a vested interest in this, we are more than happy to connect and talk.

Moderator: Thank you. That is the final pending question, but I’d like to give you the opportunity to make any concluding comments that you’d like to.

Dr. Hadi Kharrazi: Just ten seconds. I appreciate your time joining this presentation and I’d love to hear from you, your feedback, any comments, suggestions on where should we go after this. I’d love to hear more from the audience. So thank you again and wish you all a good day.

Moderator: Excellent. Well thank you so much for coming on and lending your expertise Dr. Kharrazi. We really appreciate it. And thank you to our attendees for joining us. This has been part of our PACT Demo Lab series presentation and these take place every third Wednesday of the month at noon Eastern. So the next one will be taking place on April 17th and that will be Patient Care Assessment System, PCAS, for 2019 Updates: Opening Our Door for All VA Care Providers. So once again thank you for joining us today and please fill out the feedback survey as it populates on your screen once I close out the session. And thank you once again Hadi. Have a great rest of the day.

Dr. Hadi Kharrazi: Thank you. Bye.

[ END OF AUDIO ]