Cyberseminar Transcript

Date: April 6, 2020

Series: VIReC Database and Methods Seminar

Session: WEIGHT WEIGHT…DO TELL ME!: Comparing Algorithms to Process and Clean CDW Weight Data

Presenter: Jennifer Burns, MHSA; Laura Damschroder, MPH; Richard Evans, MS; Wyndy Wiitala, 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: Hello, everyone. And welcome to Database and Methods a Cyberseminar Series hosted by VIReC the VA Information Resource Center. Thank you to CIDER and Heidi for providing technical and promotional support. Database and Methods is one of VIReCs core Cyberseminar series and it focuses on helping VA researchers access and use VA databases.

This slide shows the series schedule for the year. Sessions are typically held on the first Monday of every month at 1 p.m. Eastern. Although some dates have been adjusted due to federal holidays. More information about this series and other VIReC Cyberseminars are available on VIReCs website. And you can view past sessions on HSR&D’s VIReC seminar archive.

A quick reminder, as Heidi mentioned for those of you that are just signing on slides are available for download. This is a screenshot of the sample email that you should’ve received today before the session. In it you will find the link to download the slides.

Today’s presentation is titled; WEIGHT WEIGHT…DO TELL ME!: Comparing algorithms to process and clean CDW weight data. It will be presented by Wyndy Wiitala, Richard Evans, Jennifer Burns, and Laura Damschroder. Wyndy joined the VA Ann Arbor Center for Clinical Management Research in 2009. She currently leads a qualitative core providing support and mentorship to the center’s quantitative data staff. Her interests into quantitative methods for secondary data analysis, prediction modeling, and processes for promoting reproductivity of research. Rich is a Statistical Programmer in the VA Ann Arbor’s Center for Clinical Management Research. With an MS in Biostats, Rich specializes in R, SAS, and SQL programming with VHA data sources. He has been with CCMR since April of 2018. Jenny is a Data Manager who has been at the VA Ann Arbor’s Center for Clinical Management Research since 1993. She has many years of experience working with various VA data sources from FileMan and VistA to MedSAS files at Austin, to CDW and beyond. Laura joined the VA Ann Arbor’s Center for Clinical Management Research in 2001. Her research interests are in implementation research methods and has led several projects using weight and other CDW data; alone or in combination with data collected as part of clinical trials. Thank you so much for joining us today, enjoy the seminar.

Laura Damschroder: Hello. I guess I’ll just kick in. This is Laura Damschroder and I’m going to start us off. And Rich just to confirm you have control of the slides, is that correct?

Richard Evans: Correct!

Laura Damschroder: All right. So as was said we are going to spend this hour talking about use of weight data within the CDW database. Next slide.

So first we wanted to start off with a poll so that we can get an idea of who all’s on the phone. And we would like first to know, what is your role? Just talk about where, or put the circle where you spend most of your time whether it’s as a methodologist, whether it’s a data manager, analyst, or programmer. The third option is a project coordinator. And the fourth is other and then we’re asking you to just describe your role via the questions function if you choose other.

Heidi: And responses are coming in. I think we forgot one of the options that was available there, but that’s okay responses are coming in. If your role is other please use the questions function to send that into us. Responses are slowing down so I’m going to close this out. And what we’re seeing is 11% methodologist, 65% data manager, analyst, or programmer, 11% project coordinator, and 14% other. And the response received in the other is nursing informatics. Thank you.

Laura Damschroder: All right. Thank you. It’s great to see so many analysts and also varieties of other people as well. So we hope that the content of our presentation today will both provide details for those of you that are more familiar with CDW. And also an orientation for those of you who are less familiar. And on that note we want to go to the next poll where we’re asking you to rate your expertise in VA weight data. So using clinical weight and do you have this poll set up? Because I’m not seeing it transfer to the poll question. And if not, then I’ll just skip forward to the next slide.

Okay, let’s go to the next slide. So what we’re presenting here today is actually part of a larger study called Data-Driven Care for Evaluation for Prevention within the VA. And as part of this work we conducted a systematic literature review and we do have an article, the citation there, is there on the slide. And it is in press. It is not yet published online at least as far as we know. But do keep your eyes open for this and a lot of the details of our literature review will be in this paper. But today’s presentation is really kind of in recognition. When we started the work for this project recognizing that a lot of researchers use weight data, first of all. And secondly used a range of approaches for kind of cleaning the data, preparing the data for studies. And in the literature review we identified 492 published articles using VA weight data in some form or fashion. And through our eligibility criteria for those articles narrowed it down to 39 articles.

We want to thank our operational partners who have funded this work. The VA National Center for Health Promotion and Disease Prevention. Doctors Kim and Goldstein, Raffa, Califano, and Spohr. And just have been wonderful and supportive partners all the way through. Collaborators in Durham and Indianapolis. And then also volunteer researchers and also operations within VA who lended their expertise and feedback on early versions of this list. Next slide.

And with that, I will turn it over to Wyndy Wiitala who will provide, kind of walk us through and introduce the agenda. But our objectives for today are to compare algorithms that we identified in our literature review that are used for extracting and processing clinical weight measures from CDW. And then secondly to provide guidance and recommendations for choosing algorithms to use when using data in research and evaluation. And with that Wyndy I’ll turn it over to you.

Dr. Wyndy Wiitala: Great. Thank you so much, Laura. So I’m going to talk a little bit for a couple of minutes on the background for this work. And kind of share with you our perspective as we have been going through this work.

So as many of you are aware over the past decade there’s been a rapid increase in the use of Electronic Health Records nationally which has made vast amounts of information available for use. In the VA of course we’ve been using EHRs since the 1990s. So many of our data tables in the CDW of course contain data from this time. And so we have a lot of data available for use in research and evaluation projects in the VA. But of course there are issues associated with using Electronic Health Record data for research and evaluation. And some of these issues include the lack of control over data definitions and the data collection process. This is secondary data of course so we don’t have any control over what is collected. And then there’s methodological challenges associated with processing and transforming the raw and messy Electronic Health Record data. So therefore there have been calls for increased transparency regarding data cleaning efforts, methods to assess the EHR data quality, and increased reporting and sharing of methods.

So as Laura mentioned the work we are presenting today, we examined and compared different published methods for extracting and processing patient weight data from the CDW. So weight is a common clinical measure that’s used in many research studies. It is frequently measured resulting in many measures per patient over time. And it can also vary substantially within patient. There may be different units of measurement, may be subject to data entry errors, and all kinds of different things that may happen with the data when it’s extracted into the data tables or from the Electronic Health Record and the way it’s recorded there. And so we all know that this can be very challenging in terms of cleaning this data to use in a research project. And there’s no standard for processing and cleaning the electronic health weight data. So researchers are then left to their own to develop algorithms to define weight which then of course results in many different definitions in the published literature. And these definitions can range from simple cutoffs of implausible values or outliers to more complex algorithms that require significant coding and processing. And also may be difficult to replicate across studies.

So in particular it is unknown how the resulting weight measures may vary based on how researchers process and clean the data. And then what is the impact of algorithm choice on results and research findings. And this is also unknown.

Okay so I’m going to hand it over to Jenny Burns who will talk about the data that we used in this work.

Jennifer Burns: Thanks, Wyndy. This is Jenny Burns and I’m going to talk a little bit about the data and where exactly we extracted data from for this project. To start, this is just an overview of kind of the data collection process for this project. We’re working with the CDWWork database and we’re gathering all Veterans who had a primary care visit in 2016. Those who are at least age 18. And then we’re selecting their first primary care visit in 2016. And then of that group we’re randomly selecting 100,000 patients. And for those 100,000 patients we’re gathering all their weight from two years before and after that primary care visit.

Okay, so we’re going to go into each of those in a little bit more detail. So starting with the cohort we’re selecting patients from the outpatient workload with a visit date in 2016 with a primary or secondary Stop Code of 323. And we dropped those where the date of birth on the workload record calculated their age as less than 18. Just assuming those were mistakes. And then as I said we randomly selected 100,000 resulting patients and selected their first or earliest visit in the year. This is what we call their index date. And I just want to mention that we also pulled data from 2008 because we wanted to see if there was potentially changes in the weight data over time. But we essentially found very similar results so we decided to just report the 2016 cohort here.

Thanks. Okay so the weight data comes from the Vital.VitalSign table. Where the VitalType equals weight and the VitalSignTakenDateTime variable is between two years before and two years after the index date. This is what we call the raw weight data. I just want to mention VitalSign is at this point the only place in CDW that you can actually get weight data.

All right and we did exclude one group of patients and those were women who were pregnant during the four-year timeframe. We searched in the outpatient VDiagnosis and the inpatient InpatientDiagnosis table to exclude those patients. Obviously you know there pregnant women just have a whole different trajectory of weights than most people do. So we just completely excluded those from our samples.

One other thing we decided to flag and we’ll hear more about this later in the results, was diabetes status of patients. So we looked both two years before and two years after the index date, again in the Outpat.Vdiagnosis and the Inpat.InpatientDiagnosis tables. Patients were required to have two or more outpatient or one inpatient diagnosis to be considered diabetic. And that’s kind of a standard algorithm that we’ve used in a lot of research for defining diabetes.

And now I’m going to turn it over to Rich Evans who’s going to talk about the algorithms and then go through the methods and results of our study.

Richard Evans: Thank you, Jenny. This is Rich Evans. I’m going to talk a little bit about algorithms. So we talked already about the systematic literature review, it’s in press. Of that, from that systematic literature review we used, we looked at studies that utilized patient weight outcome measures or patient weight in general from the VHA CDW databases. And then of those, well actually there was 430 or something studies that we looked through, 39 of those there was a least of CDW to define patient weight outcomes. Of those 39 we found 33 that included a weight cleaning algorithm that we could implement and replicate based on their methods, the published methods.

So we present here 12 representative algorithms and we’re going to provide details about the remaining algorithms in the supplement and within our GitHub repository.

So of the 12 algorithms and I should mention now these 12 algorithms were chosen because they are of the 33, they were the ones that were most representative and nonredundant. So we divided these 12 algorithms into two conceptual groups, mostly equal groups. Those that include all weight measurements during a specified timeframe. And those that were time-period specific. So the time-period specific algorithms selected maybe as an example baseline six-month and 12-month time periods. And included weight measurements during some kind of window around each of those time points, in 30 days. And note that not all algorithms fit into these groups.

So just a high-level overview of the main exclusions that occur after applying each algorithm or what was involved in each algorithm. Split into the two groups that I had mentioned; one that used all weight measures and ones that are more time-period specific, on the right. We just choose a few of them. You see them, you view it on the left. Most of them actually, well let’s back up. Most all of these describe a method of first removing outliers based on some implausible value cut. And then some of them go further in-depth. So we view in 2013 what they do is they incorporate a linear mixed model for each person. And then use the absolute value of the conditional residual greater than 10 as ones that we think are implausible and removes those. And therefore the result in weight or outcome or output is cleaned weight. If you look at on the other side, time-period specific. And look down to Noel 2012 same thing, cutoffs, simple cutoffs. But then patients with too few values to compute the median within fiscal quarters. So they divided the time into fiscal quarters and then they take the median from each one of those. And a lot of them kind of do something kind of similar to that.

So just a brief jut into our GitHub. If you were to go to the CCMRcodes GitHub and we’ve published all these algorithms in this work to this GitHub repository. You go to the repository circled in red there. Look for the weight algorithms one. And then if you look down into, you click on that, you’ll find ClinicalCodes that we used to define certain things. But our programming which defines how each of the algorithms were actually coded up. And then sometimes there’s SAS code for each of those and then SQL code describes how we actually grabbed these data.

And if you keep going there you’ll see the review in one. You can click on that and you can download each of these algorithms yourself, play around with them, with any data that you’d like to have, evaluate.

So now we’re going to go through some methods and results. And we chose five different scenarios common to our research questions, research implementations that we see in the literature all the time. Descriptives, using weight as predictor, weight change between one timepoint and another. Weight trajectories and facility-level measures or performance measures.

So as I just said, so as a predictor in studies it can be seen as something that seeks to adjust for the effect of baseline weight when examining the association between another variable and an outcome. And we look at weight change which is in studies examining the effect of an independent variable on patient weight or weight change over time. And then we look at patient weight trajectories which is kind of a longitudinal analysis method. And then facility-level effects. So maybe in studies that are examining performance measures across facilities, groups, or other clusters.

First up is descriptives.

So what we did was all algorithms that we programmed were applied to the data, the raw data and for each of the two cohorts were really wealthy for now just the PCP, the 2016 cohort. Then we compared based on the descriptive statistics including the number of measures the patients retained, the mean, standard deviation, and variance and all that stuff. For comparison we also included descriptives based on the raw unprocessed weight data during that same time.

So if you look at the highlighted column there. The number of patients retained is largely similar across almost all. Almost in the 90s for all except for an outlier there would be Kazerooni down at the bottom, 24%. Rosenberger maybe 64%.

The number of weights retained is very different between number, the ones that utilized all data and then ones that are time-period specific. That makes sense. So time-period specific right around 15%, almost 20% are retained.

But despite all that no matter how you choose it, because we have a large amount of data that we collected, the mean in the standard deviation of variance do not change much. Hovering around 200, 205, and maybe 45, 46, 47 for the standard deviation.

However even the slight, having an algorithm that cleans your sorely left with some values that are quite implausible 1,000 pounds, 1,233 pounds, 2,423 pounds in the raw weight data.

And key takeaways here is that despite cleaning efforts implausible weights do remain in the data. And for large cohorts of patients the loss of data due to the algorithm choice does not appreciably change the mean and variance.

Now we’re going to move onto weight as a predictor.

To compare algorithms in this context we presented an example that shows the association between baseline weight and new-onset diabetes. To do this we excluded patients with diabetes prior to that study index date. Then we defined new-onset diabetes as the presence of two or more diabetes diagnosis codes after the patients’ index date. And then we applied each of the 12 algorithms to describe baseline weight measures for the patients in our four cohorts; one cohort for now using weight measurements that occurred during a 60-day window on or before that index date. The resulting baseline weight measure was the measurement that occurred on the closest day to the index. And then we used logistic regression models for each of the algorithms to obtain an odds ratio and a confidence interval for the effect of patient weight on new-onset diabetes. We’re going to show a graphical [unintelligible 23:35].

So we look at this scale. It’s a very small change in odd ratios between each. You might be able to see a small outlier in Kazerooni but it’s a very small percentage change of odds ratio. Depending, and it doesn’t matter which one you choose.

But basically the key takeaway here is that there are relatively small differences in the effect size. And across algorithms the CIs overlap substantially. And some of the time-period specific algorithms just show some greater variation.

Now we’re at weight change.

And these are weight loss greater than 5% of some baseline weight. And then we similarly looked at weight gain greater than 5%. And to compare each of these we applied each algorithm to the cohort. So then we used a 60-day window to define initial weight values and we included a weight measurement taken on the closest day to the index. To define one-year follow-up weights we again used a 60-day window around the date one year after baseline. And keeping the closest weight measurement.

What you see is that a number of patients retained. It’s a little bit higher, about 20% more maybe in ones that utilized all data. The ones that are time-period specific can show drastic losses of patients. So Kazerooni, Rosenberger.

And then the weight loss despite that is relatively similar. So you’re looking at 13%, 12% across, removing an outlier from Maguen of 9.4%.

And now weight gain greater than 5% still is steady, stable between each of the algorithm choices.

One thing that’s a takeaway from this is that with the Rosenberger, basically the min and max can still, despite all the cleaning that has been done here and all the processing there are still outliers that are implausible when it comes to weight change.

So weight gain/loss proportions are fairly consistent across algorithms. But implausible weight values still remain even after applying sometimes a possibly complicated algorithm.

We’re now going to look at weight trajectories.

So longitudinal weight trajectories are also called group-based longitudinal modeling. And so researchers may be interested in assessing weight trajectories within patients over time. Potentially classifying patients according to their trajectory or examining whether types of patients respond differentially to some sort of intervention. And we know that algorithm choice may impact the trajectory of individuals and their measurements collected over time. Especially for algorithms that might reduce the number of measurements that are left to analyze. So instead of aggregating patient weight over a specific time period which would be in the last session, studies analyzing weight measures utilize repeated measures designs may be generalized linear models, mixed models, ANOVA/ANCOVAS repeated measure ANOVA. So to compare algorithms in this context use a latent class mixed model that assumes the population is heterogeneous and composed of some selected latent classes characterized by specific trajectories.

Now we’re going to go through a small vignette to show what this might mean. So if you take the raw weights and use the latent class mixed model what you get is three different predicted latent class. There’s a gain, there’s a loss, and a maintenance group.

And then it’s the same from before but adding another algorithm outputs of Breland 2017. So it shows a gain with a parallel slope but a very similar weight loss and maintenance group.

Adding another one you see a very large difference with the Kazerooni one. It looks like much higher weight gain but very similar maintenance group. And a loss group that might not actually be very different.

Well let’s put them all together and it’s the same colors as before, the same algorithms. If you put them together and you see vast differences. And so this could impact your estimation and research output depending on how you actually clean.

So algorithms that use all data to clean weight measurements appear to be more appropriate when analyzing longitudinal trajectories.

Now we’ll look at facility-level.

Facility-level measures we use the space similar thing that we did with the weight change outcome. So researchers and evaluators are often interested in comparing facilities according the percent of patients that meet this metric of interest. So we used the raw data and each of the 12 algorithms to calculate the percent of patients at each facility with either one-year weight loss greater than 5% or one-year weight gain. And then our objective was to understand the impact of the algorithm choice on calculated facility-level metrics therefore we examined unadjusted facility rates. We unadjusted for any kind of measurement. Then we ran ordered those facilities based on the percent of patients meeting each metric. And then what we’re going to show you is that we compared the differences in the facility-level percent of patients based on each algorithm. And then within each of the ones that used all data and time-period specific algorithms.

And that’s this, right here. This is weight loss. What you see in the time-period specific is that there are, there’s variation. There’s a lot more variation compared to the right, the use all data. so there’s a black line there in the middle and that’s raw data across facilities which I should mention is not a gold standard. So this is just differences from some raw weight data and how you change it. There’s one algorithm there. It’s kind of hard to tell but it shows vast differences in the each, for, between each of the facilities. Look at all use all data one of the algorithms shows vast differences apart from all the other algorithms. In fact it’s the only one that really shows any difference.

Weight gain is almost the exact same picture.

And key takeaways here, with some exceptions algorithms that use all data seem to exhibit less variation in measurement.

So here’s some findings and recommendations.

Our principal findings. The differences between algorithms are quite minor implying that for many studies a simpler algorithm design may be computationally more efficient. And that in some cases the results are not different than using raw, unprocessed data despite algorithm complexity.

So what we recommend. The studies using point estimates of weight and weight change may benefit from simple cleaning rules just based on those outlier cutoffs what we would deem implausible values which might be subjective. And for trajectory analyses or longitudinal thing time-period specific algorithms may not be appropriate meaning that the ones that utilize all data are better estimated. For facility-level measures all time-period specific algorithms result in inconsistent results compared to algorithms that use all data.

So what we recommend including, we also recommend in general from this work is that including detailed information on how measures are constructed in publications and people should share code via open source repositories.

So limitations, things that we’ve had some issues with or will become apparent. Is that the algorithms were reconstructed from published methods. So there’s potential for misinterpretation. We only reached out to few authors to kind of get an idea of what things might be missing. Mostly there was enough information in the methods to accurately, according to us, reconstruct one of these algorithms. Another is that there is a lack of gold standard. We used raw data and [unintelligible 32:11] from the raw data as sort of a measurement of these algorithms. But when we have a gold standard this might be [inaudible 32:20]. So we use VA data which obviously is non-intervention/clinical sample. We use a sample and it’s a very large sample. So things will change quite a bit when you choose smaller samples or when you’re looking at facilities which are, which can have small cluster sizes and large cluster sizes.

So to follow-up with that, for works in progress we’re working on a simulation study that will address that lack of a gold standard and large and small sample size.

All right. And then we have another poll.

Heidi: Sorry about that. I did get the poll open this time. How useful are these recommendations in your work? Responses are coming in we’ll give everyone a few more moments to respond before we close the poll out and go through the results here. Sorry that I wasn’t able to get the last poll to open. My computer completely froze up, Rob was trying to do it and he just wasn’t able to get to it quick enough. But, we got it this time. Okay looks like we’ve slowed down here so I’m going to close this out. And what we’re seeing is 42% of the audience saying very, 42% moderately, 15% slightly, and zero not at all useful. And we did get a comment in from one person who answered slightly because they don’t deal with a ton of data currently. Thank you, everyone.

Richard Evans: So there’s my face. Thank you. You can send any additional comments or questions about the data, the algorithms directly to my email account. And anything on the GitHub can be addressed to me as well. Also, yeah this is a, I think we’re done here. I think we can do questions and answers at this point.

Laura Damschroder: Yeah and this is Laura, we went through the results pretty quickly. There is a lot to absorb. As was said earlier the slides are available online. Feel free please to enter questions in the chat or the question function within the software. And if there are any slides you’d like to go back to, to kind of take a closer look be able to absorb the information, we would be happy to do that we’ve got some time. Wyndy, I don’t know if you have any other information to ask.

Dr. Wyndy Wiitala: No, I think that’s right, Laura. We’re happy to take additional questions and focus on what might be useful for the people on the call today. And I think the other thing that Rich also mentioned that I of course want to reiterate is just that we have made available all of this code on our GitHub page and so we would love for people to check that out as well. And share with us you know what they’re doing and/or questions, feedback, comments, all of that even after today’s seminar.

Moderator: And we did have a question regarding the GitHub. Is that a VA GitHub or GitHub.com?

Dr. Wyndy Wiitala: It’s a [unintelligible 36:15].

Richard Evans: GitHub.com.

Dr. Wyndy Wiitala: Yes.

Moderator: And then we do have some other questions coming in. Are all weights entered in the same unit and are units available?

Dr. Wyndy Wiitala: Jenny, do you want to take that one?

Jennifer Burns: Sure. The units are not available. The weights are supposed to be entered in pounds but we know that that’s not always the case. I think part of the reason for the sort of min and max cutoffs of less than 75 and the greater than 700 or whatever kinds of things are usually to get rid of things that are probably kilograms. But there could be errors in there still. And hopefully some of these cleaning algorithms will find that.

Moderator: And we have another question. Can you go back and explain what exactly all data means versus having a window?

Richard Evans: Sure. So when we describe an algorithm that uses all data, let me go back to a good slide for that. All right. So when we look at algorithms that use all data we’re looking at algorithms that in order to estimate something they’re going to use all remaining weights that are left after some sort of cutoff or some sort of processing. So they’re not choosing baseline, six months, 12 months, or something like that. So it would be something like maybe you have a person that has 200 weight values collected over maybe a 4-year period. But you’re only choosing three for something that you’re looking at. So these ones might actually be able to use all weights that are collected in order to make some sort of measurement of weight. So you get a bit of, more of, a better picture of that person’s weight. So what we mean by windows let’s say you selected I want the weight at baseline. And then, but the thing is there might not be a weight there. So window would be 30 days surrounding that day. So maybe you’re collecting, if there’s a baseline weight at zero, minus 30 days or plus 30 days you collect whatever that is. So it’s kind of a fuzzy definition.

Moderator: Great. Another question is do the algorithms convert potential kilogram weights into pounds?

Richard Evans: None of them really do that with, at least explicitly as defined in methods. That’s because most of these work, I mean they were all taken from VHA CDW and as Jenny said we don’t know the units. And so as we were saying that the cutoffs kind of guard against improper, basically kilograms versions of the weights.

Moderator: Okay. And is there any consideration of amputation and its effects on measuring weights?

Richard Evans: We have not seen any algorithms that deal with that yet. That is something that we’ve been thinking about.

Moderator: Okay.

Laura Damschroder: Yeah I think one of the things to keep in mind, this is Laura, is that we did limit the algorithms that were applied specifically to secondary data. To you know see, in this case the EHR data is captured in CDW. A lot of times we see you know more focus on imputations or approaches to address missing data in a clinical trial context where you know assessments are, you know there’s a primary data collection I guess specific to the trial following a specific measurement protocol. At least ideally. And in doing an intention to treat analysis whereas this approach and Wyndy I don’t know if you, if I’m on the right track here but you know with this data we’re using the data and trying to really optimize use of data that are already available.

Dr. Wyndy Wiitala: Hi, so this is Wyndy and yes that’s right. I also want to just mention the question about the kilograms versus pounds. I think that for some of the algorithms that we’ve classified as those that fall under using all weight measures. I think that some of those algorithms will get at this a little bit more. Particularly in the event that one patient has weight measured both as pounds and kilograms and it would, and sometimes we see this right that a patient maybe has six weight measures and one of them is in kilograms and the rest are in pounds. And that kilogram weight measure is a clear outlier for that patient. So potentially some of the use all weight measure algorithms might identify those weights that might be an outlier within patient as opposed to sort of like the global cutoffs. Which also would help us to get rid of some kilogram measures too.

Moderator: And then going back to the previous question, it seems like there might have been a misunderstanding. They were asking specifically about amputation and weight measurements.

Laura Damschroder: Oh with an A.

Moderator: Yeah.

Laura Damschroder: For patients with amputations?

Moderator: Amputations.

Laura Damschroder: Oh.

Moderator: Sorry I have a bit of a cold. It was probably on me.

Laura Damschroder: That’s all right. I was thinking you know we immediately go to imputations\_

Moderator: Right.

Laura Damschroder: \_from a statistical [unintelligible 43:01]. So maybe that goes to, I mean we didn’t look for flags for amputations so there may be patients with amputations in the database. I don’t know if you have any reflections like how to identify or?

Jennifer Burns: No, actually it’s something we hadn’t thought of. We did, we looked at patients who had bariatric surgery and we also flagged weights that were from inpatient stays. And so we wanted to do some potentially sensitivity analyses with those. Where we take either those patients that had bariatric surgery out or the inpatient weights out. But we haven’t done that part yet. But I think doing that with amputation patients actually might be a really good idea.

Richard Evans: Yeah and I don’t believe that we’ve seen any algorithms that specifically deal with amputation or any studies that used it, an algorithm in the context of amputation.

Moderator: And then, how did you decide on the five different methods for evaluating weights?

Richard Evans: These were, all five of these were things that we’ve seen in the literature before. So descriptives are very common across all studies, longitudinal trajectories are very important in weight modeling or help interventions, weight interventions. Weight change is another one that comes up in two different places. When it comes to just individual weight change [unintelligible 44:51] measure. And then as a predictor that, yeah that comes up a lot I believe.

Moderator: Great. And I know several people are asking for the GitHub URL so we can send that information afterwards.

Dr. Wyndy Wiitala: And it is included in the slides as well.

Moderator: Oh, okay.

Dr. Wyndy Wiitala: Yeah.

Richard Evans: Yeah so it’s right here. You just go to GitHub.com/CCMRcodes and then if you were to just another slash, weight algorithms you would find it.

Moderator: Wonderful.

Richard Evans: But, yeah it’s in the slides here.

Moderator: Great. And I’ve heard that sometimes data cleaning could impact subpopulations of patients differently. Did you look at that or consider that?

Dr. Wyndy Wiitala: So Rich maybe you can respond to this too. But we haven’t looked at that explicitly. We have some plans to do that. We thought about that in terms of looking at subsets of patients with different diagnoses, looking at other sort of demographics, and things like that. But we have not done that yet. And with the exception of the ones that Jenny had mentioned in that of course we excluded any patients who had pregnancy codes entirely from the analysis. But then also we did flag patients for bariatric surgery and things like that so we could potentially look at some of those subgroups.

Moderator: And how do you plan to use your findings?

Richard Evans: Well one of them we’ve already tried to, well we’ve already used in a tangential project called, we’re looking at CoAlgorithm implementation across facilities. Using at least one of these algorithms for applying that and then cleaning weight data. But there are many others. I don’t know if anyone else wants to chime in.

Dr. Wyndy Wiitala: So I think this was a really interesting study for us in that we kind of anticipated from the outside that we might see more differences. You know since some other kinds of things in what we saw. And so I think that it was kind of interesting to us that you know depending on our purpose for using the data, depending potentially on our sample size but also then for depending on the type of analysis that we’re doing, that that might sort of help guide our decisions in terms of the amount of time that we spend making sure the data is super, super clean. It was kind of unknown from us, for us from the outside. We’ve all, in this project had all used weight data previously and always kind of wondered well where do we stop? How much cleaning do we do? Where do we stop? Because you know it’s really challenging to clean it entirely. You know you’re always going to have some error but you’re just not sure you know where to stop and how much effort to put into writing some complex code to clean the data. And so that was really informative to us and I think going forward that we’ll think about the type of analysis we’re doing and help, have to have that sort of guide us in terms of the extent to which we spend time and processing and coding in order to clean the data.

Moderator: And have you evaluated how quickly or efficiently each algorithm completes its cleaning.

Dr. Wyndy Wiitala: So I’ll let Rich really respond to this. But we did think about that and looked at it. And one of the issues is, is that you know there’s a lot of things when we access CDW data right and we access things that are sort of beyond our control, right. And so there could be a lot of variability in terms of the amount of times to process just based on things that we’re unaware of. You know the traffic on the server and things like that. So that becomes a little challenging. But yeah, definitely I mean and if you look at the code for the algorithm sets some of them are quite complex. Estimating you know a linear mixed model whereas others are just saying okay we’ll let’s just drop weights outside of this weight and we’re going to take weights within this timeframe. So those are a little bit quicker. But Rich I know you’ve thought about this and looked at it a little bit earlier on.

Richard Evans: Yeah so basically it’s a question about benchmarking algorithms. We know which ones are the slowest. There might be ways to clean them up, to make them a little bit faster. But that wasn’t really what we were doing right now. When it comes to getting these algorithms out there and like try and make them a little bit quicker. But yeah, we didn’t really look at that. But the ones that use like a mixed model on large datasets is very, is going to be very slow. And there’s another one that I believe that’s comparably slow. But it just becomes, it's just that way with larger uses, sets of data.

Laura Damschroder: And it also, I think that you know part of the consideration is just which kind of hints that is cognitive I guess efficiency let’s just say. You know simpler is easier to describe. Maybe easier to you know defend and provide rationale. And more complex algorithms are more, you know of course more difficult to describe and more difficult for people to understand. And also to replicate in their own work.

Moderator: Great. And we have lots of positive feedback. People are, they love this work and very excited and think it’ll be very helpful for researchers as well. So I just want to make sure that you guys also have that feedback.

Dr. Wyndy Wiitala: That’s great, thank you.

Richard Evans: Thank you.

Laura Damschroder: Yeah, thank you. And just want to reiterate that Rich had his email up there, feel welcome to email any of us and we will get you to the right person. We definitely, we love working with this data. We’re kind of geeks about it. So appreciate the enthusiasm and do appreciate you know any questions or insights anyone else has to offer.

Moderator: Okay, well it looks like that’s the last of the questions coming in right now. But obviously as Laura said you can email them questions afterwards. So, we can switch over. I wanted to thank Wyndy, Richard, Jennifer, and Laura. Thank you so much for taking your time to present today’s session. And to the audience if your questions weren’t addressed during the presentation you can contact the presenter directly. And you can also email VIReC HelpDesk at [virec@va.gov](mailto:virec@va.gov).

And then, oops, and then we have our upcoming Cyberseminars. So please stay tuned for the next session in VIReCs Database and Methods Cyberseminar Series on Monday, April 6th [sic] at 1 p.m. Eastern. Dr. Charles Maynard will be here to present Ascertaining Veterans’ Vital Status: Data Sources for Mortality Ascertainment and Cause of Death. We hope you will join us. And thank you once again for attending. We will be posting the evaluation shortly. Please take a minute to answer those questions, let us know if there are any data topics you’re interested in and we’ll do our best to include those in future sessions. Thank you, everyone and have a great day.

[ END OF AUDIO ]