Heidi Schlueter: And once again, thank you, everyone, for joining us today; we are going to go ahead and get started. I would like to start off by introducing today's presenter. Our presenter today is Arin Madenci. He completed his PhD in Epidemiology at Harvard School of Public Health and is a resident in general surgery at Brigham and Women's Hospital in Boston. Through work with Professor Miguel Hernán at Harvard, he has developed an interest in adapting contemporary epidemiology methods to important research questions in clinical medicine. He has been fortunate to have had the opportunity to work with the Massachusetts Veterans Epidemiology Research and Information Center, MAVERIC, over the past two years and looks forward to continuing his collaboration.

Before we turn things over to Dr. Madenci, we are going to Mihaela Aslan, the Deputy Director of CERC, will be doing a quick introduction for us. Dr. Aslan, can I turn things over to you?

Mihaela Aslan: Absolutely. Thank you, Heidi. And good afternoon and good morning, everyone. I am Mihaela Aslan, and I'm the Acting Director of the West Haven Cooperative Studies Program Epidemiology Center, and I would like to take just five minutes of your time on behalf of the VA Causal team, to mention that Arin's presentation today was made possible by the larger VA Causal Inference Enterprise, an ORD-led initiative by Dr. Ramoni, and sponsored through the Cooperative Studies Program.

In terms of background and objectives of the VA Causal Program, as I mentioned, it's and ORD-sponsored initiative and it's part of the VA Research Enterprise who are aiming to integrate causal inference research across the VA with the objective to establish a VA learning healthcare system by identifying and implementing effective and safe clinical interventions through a decision-making process based on both high-quality data generated by the VA system itself, as the largest healthcare system in the US, and causal influence methodology implemented at scale to improve veterans' health.

The program structure related to the executive committee primarily was established on November 26, 2019. I have here listed the executive committee members in alphabetical order; there here are two cores of the program: the Methods Core, which is led by the CSP MAVERIC at Boston VA and the Harvard T.H. Chan School of Public

Health; and an Implementation Core, which is led by the CSP West Haven Clinical Epidemiology Research Center and the CSP Boston Clinical Trial Coordinating Center, with the West Haven CSP Clinical Epidemiology Research Center being the coordinating center of the program.

Here, we have a very nice summary diagram courtesy of Professor Hernán. As you can see in just one picture, we can see that VA data in the form of you name it mvp biobank CDW and also CSP randomized trials will be used for the VA Causal engine that plans to apply causal inference methods, not only to identify therapeutic targets, but also evaluate and confirm those results of effectiveness and safety through experimental constraints and, of course, extending them to the larger VA population with the ultimate goal of implementing generalizable results at point of care and improving veterans' health.

Currently, we are working--both the Boston and the West Haven team--are working on six use cases; Arin will be presenting on the very first one that's across the finish line. Just like bariatric surgery, number four is another single-point intervention use case that will be applying similar methods with the other four listed here having a little bit more complicated designs.

In terms of immediate next steps, we would like to, through this webinar, to have a first step of disseminating information; but ultimately, we would like to have the VA-wide community of researchers adopt the tools and methods created through this program. So, right now, we are working to identify a centralized location for placement of causal influence resources, but also, we would like to ask all of you to really reach out to us with any other single-point intervention use cases you have in mind or you are working on and would be interested to apply these methods. We are also creating a director of the VA investigative users to really enlarge our community of causal inference investigators.

You can reach out to any of us presenting today; you can also reach for general questions and contact information by emailing Patricia Crutchfield, who is the program manager of the VA Casual and will be very happy to hear from you. So, at this point, I would like to turn it over back to Arin.

Arin Madenci: Thank you very much, Heidi, for that introduction. And to Mihaela for introducing VA Causal. Thanks very much. The title of my presentation today is Estimating the Effects of Bariatric Surgery on Cardiovascular Events Among US Veterans. This is a preliminary analysis of one of the initial works with the VA Causal Enterprise. I don't have any disclosures.

Before getting started I would like to make a few additional acknowledgments. One is to the MAVERIC group including JP Casas, Kelly Cho, David Gagnon, Mike Gaziano, and Katherine Kurgansky, all of whom have been incredibly helpful and supportive in, one, introducing me to the vast resources that the Veteran Affairs Hospital Enterprise has and assisting with research. And next, at the Harvard T.H. Chan School of Public Health, Barbara Dickerman and my advisor

Miguel Hernán. And finally, in the VA Boston Healthcare System, one of the bariatric surgeons, Dr. Ann Smith, works at Roxbury VA Medical Center and has been integral and helping to guide this work; and then my residency program at Brigham and Women's Hospital in the Department of Surgery.

The goals for this presentation will be to outline this pilot project. First, we'll describe the methods, we specify the hypothetical randomized trial that we would like to conduct if we were able to--and this is called the target trial. Next, we emulate that target trial using observational data; and then I will go ahead and present our preliminary results. I will share some background details related to our substantive area of backgrounds of bariatric surgery and then introduce our specific causal question.

The word "bariatric" comes from or is derived from the Greek "város" or "weight"; and then "iātrik", "treatment of", meaning weight loss surgery. This cartoon demonstrates the three most common bariatric surgeries performed including a gastric sleeve, which is, perhaps, the common; the lap-band, which is becoming less and less common; and gastric bypass. I'll highlight that two of these surgeries are endocrine surgeries: gastric sleeve and gastric bypass; the lap-band is a weight loss surgery, but it does not decrease the absorptive surface area in the same way as a gastric sleeve and bypass

One thing I would like to highlight about these surgeries that is relevant for our target trial, is that there is an extensive process that occurs between the time a patient is referred to a bariatric surgeon, or an overall weight loss program, and the time of surgery itself.

And this is an example of a pre-operative checklist when a patient is injured is interested in bariatric surgery. This includes completing a routine history and physical exam, laboratory values, nutrient screening, full cardiopulmonary evaluation, including obstructive sleep apnea screening, and cardiac disease; a GI evaluation, endocrine evaluation including diabetes or hyperthyroidism, and other endocrine disorders; lifestyle evaluation and interventions including fitness, strength training, sleep hygiene, substance use, and so on. This is the second part of the preoperative checklist involving an RD for nutrition evaluation, psychosocial evaluation, and optimizing several features including weight loss and glycemic index prior to an operation; smoking cessation counseling and fast cancer screening.

As such, the median duration from initial bariatric surgery clinic date to the operation in 2016--and these are national US data--was approximately 160 days. I'll present some results of prior studies to contextualize our methods and findings. This is a randomized trial in which participants were randomized to bariatric surgery plus intensive

medical therapy or intensive medical therapy alone, and these were all individuals with Type 2 diabetes and BMI between 27 and 43. As you can see, the hemoglobin A1c is reflective of long-term--glycemic control was reduced and significantly better in the bariatric surgery arms as opposed to the intensive medical therapy-alone arms in terms of decreasing hyperglycemia.

So, this is clear evidence that the effect of bariatric surgery on diabetes is to reduce the hemoglobin A1c and give better glycemic control. However, no randomized trial has been conducted to study the effect of bariatric surgery on cardiovascular disease; a number of observational studies, all using propensity score-matching methods, have been done on this subject. I'll share some of these observational studies on cardiovascular disease. This is a study from 2012 in which the dark blue line was non-surgical patients and the gold line was bariatric surgery patients; as you can see with the y-axis being a proportion of coronary artery disease between zero and seven years after surgery; those who underwent bariatric surgery had several-fold less proportion of cardiovascular disease--coronary artery disease. Similarly, this is a Swedish group with a registry of bariatric surgery patients; and over 18 years of follow-up with a national registry, the hazard ratio for cardiovascular events was 0.8.

This is a more recent study in 2019; again, with the dark blue line non-surgical individuals and the blue line bariatric surgery, we can see that the non-surgical individuals had nearly 30 percent cardiovascular disease, and eight years in the gold bariatric surgery had less than 20 percent at eight years.

And finally, this is a Canadian study from last year in which the y-axis is cardiac mortality and we can see that there was also a significant decrease in cardiac mortality associated with bariatric surgery compared with non-surgical individuals, with a hazard ratio of 0.14.

With that in mind, we wanted to specify the hypothetical randomized trial that we would want to conduct had we sufficient resources and willingness of participants. We will then go on to emulate that trial and present the preliminary results. The eligibility criteria we would use are very similar to those of the randomized trial that investigated the effect of bariatric surgery on glycemic control among diabetics. We would include individuals, perhaps, who were under age 65 during the years 2007 and 2019; we would assess their maximum BMI in the past year. If it's above 40 and have a current BMI above 35--or if it was above 40 in the past year, but their current BMI is between 30 and 35 per the National Bariatric Surgery Guidelines, that they must also have at least one obesity-related comorbidity including hypertension, osteoarthritis, obstructive sleep apnea, dyslipidemia. And if it's between 30 and 35,

they must also have one of those comorbidities to be eligible for surgery. And then if their maximum BMI in the past year is between 35 and 40 and their current BMI is above 30, then they also must have one such obesity-related comorbidity--and this is to include patients who, at the time of clinic evaluation, had a BMI above 35 and comorbidity and were eligible for surgery, but by the time of the operation, their BMI was below 35 and they would typically still undergo a surgery in the way that the current practice patterns are.

We would, of course, want our patients not to have any contraindications to surgery. This list here that is footnoted is somewhat stricter, reflective of the randomized trial that we would want to conduct; I'll highlight that often that, for example, antiplatelet or anticoagulant medications would be able to be accommodated by perioperative management as well as several other of these depending on the severity; but we included liver failure, cirrhosis, portal hypertension, hepatitis B, psychosis malignancy, spinal cord injury, and antiplatelet and antibiotic use.

For the purpose of this trial, we ask that eligible participants have no prior cardiovascular events and we want them to be following within the VA system within the one to two years prior.

Finally, we would want them to have recent measurements and no major derangement of both laboratory values and vital signs, with laboratory values including a lipid panel, creatinine, and hematocrit, and vital signs including blood pressure and BMI.

With each step, in which I specify the way in which we would conduct our target trial or the hypothetical randomized trial that we would like to conduct, I will describe, in these green blocks, how we would emulate this step. So, to emulate these eligibility criteria, we would use the same criteria and apply them based on information from the clinical data that we have access to. Comorbidities, as mentioned on the previous slide, will be obtained from ICD-9 and ICD-10 codes, medication history, was determined from the pharmacological documentation and BMI, and other laboratory and vital sign information was recorded from the vital sign another clinical data report.

In the target trial, we might have two treatment strategies, and this reflects also the randomized trial for diabetes that was previously mentioned. So, we would randomize participants who were eligible for bariatric surgery to be performed within six months or to not undergo bariatric surgery within six months; we would do the same thing in our emulation of the target trial.

In terms of treatment assignment, in the target trial, we would simply randomly assign participants to one of those two treatment strategies;

this would be unblinded given that patients would know if they underwent a surgery. In the emulation, we're forced to do something slightly different because we do not know treatment assignment at baseline; thus, we assign all participants to both treatment groups at baseline using a copy of each participant for a clone and assigning one of two clones to each treatment group at baseline. And I will describe how we deal with the complexities of that in the analysis section.

In terms of the outcome definition for the target trial, we define this as a cardiovascular event including ischemic stroke, transient ischemic attack, or coronary artery disease diagnosis; we would do the same in the emulation of the target trial; for follow-up period, this begins at the time of treatment assignment, it ends at eight years, which was arbitrarily chosen to reflect some of the previous observational studies--it could be longer or shorter--the occurrence of a cardiovascular event, or the administrative end of follow-up which was set at January 1, 2020; and we do the same for our target trial emulation using the observed data.

Because of contrast of interest in the randomized trial, in the target trial, it could be either an intent-to-treat effect or per-protocol effects. Again, the intent-to-treat effect will be the effect of being assigned to one treatment arm or other regardless of what course was ultimately taken; the per-protocol effect would be the effect of, in fact, adhering to the treatment arm to which one was assigned. In the emulation of the target file, we consider the observational analog of the protocol effect; now, we're unable to emulate a trial in which we can estimate the intention-to-treat effect because we do not know to which arm individuals were assigned at baseline.

In terms of the analysis plan, for the target trial, we might partition a follow-up into months for each participant, and then censor into the participants at a time of deviation from their assigned strategy for the protocol effect; and then weight participants by their adherence weight. In the No Bariatric Surgery arm, this is an equation demonstrating how we would use unstabilized adherence weights; and this is a complicated-looking expression that really just means that this is the product over time of one/the probability of not receiving treatment, meaning adhering to the No Bariatric Surgery arm given the previous treatment and covariate history and not having previously experienced the outcome. And this is essentially a way to adjust for overheads may be associated with both adherence and the outcome over time. We do this over the first six months which is our grace period during which individuals are asked to not undergo a bariatric surgery.

In the bariatric surgery arm, this is a similar expression and except that over each of the first six months of the grace period, now we are interested in the inverse probability of, in fact, undergoing a bariatric

surgery given that someone has not undergone a bariatric surgery previously and their covariate history. We equally partition the number of surgeries that occur over those first six months, and we simulate a distribution in which individuals in our trial undergo surgeries in equal proportions during each of the first six months--and that's an arbitrary choice that does not change much but makes our trial more realistic.

In order to implement this, the institute pulled logistic models to estimate those previous expressions and in the vector, L, of covariates, as I mentioned there, these probability quantities are conditional on covariate patterns, we included demographics--meaning gender, ethnicity, race, age, and calendar time of entry, obesity-related comorbidities as previously mentioned, diabetes, dyslipidemia, hypertension, osteoarthritis, and sleep apnea, vital signs, body mass index and blood pressure, laboratory value results, and then prescriptions--beta-blockers, calcium channel blockers, antihypertensives, anti-lipids, respiratory medications, hypoglycemics, and insulin. In the footnote, we mentioned that for analyses limited to diabetic individuals only for subgroup analyses, we also include years since first diabetes diagnosis and hemoglobin A1c.

For the emulation of the target trial, again per-protocol effect, we make two modifications to this analytical plan. One, as I mentioned before, treatment assignment at baseline is unknown; thus, for each eligible individual, we have them contribute two clones at each time of eligibility: one is assigned to bariatric surgery; one is assigned to No Bariatric Surgery; they are censored at the time point in which they deviate from the treatment that we have assigned that clone. Additionally, we emulate what are called nested sequential target trials in order to increase precision, and what this means is that rather than only taking the first time point at which individuals are eligible, we allow them to participate at every time point of eligibility; and what this does is allows us to include more individuals when we want bariatric surgery because typically, they would not have undergone bariatric surgery just by chance at the time that they were first eligible. So, this allows us to increase the precision of the effect estimates.

At this point, I will start to show estimates from the VA data and share our preliminary results. So, this is our flow chart. At first, I'd like to highlight two aspects of this: between 2007 and 2019, there were over 5000 individuals who underwent bariatric surgery before applying eligibility criteria; I embossed that there was a random sample of 100,000 individuals taken from the overall VA population who had a maximum BMI above 35 between the years 2005 and 2019, with that extra two-year lag being present to accommodate our eligibility criteria. And this corresponded to 8,460,000 potential person-trials.

I point out that this random sample of 100,000 individuals is important because we are not able to include all individuals as we would in a true randomized trial because of the limitations of computational limitations, essentially and so that's one thing that may be worth discussion at the end and one slight deviation from our target trial.

After applying the exclusion criteria that I mentioned previously, we exclude most of the person-trials--and again, this is each eligible time point for each participant--and we are left with nearly 30,000 individuals and approximately 770,000 eligible person-trials. From those, we again had to compromise for the sake of computational capabilities, and we randomly sampled 50,000 person-trials of individuals other than those who underwent bariatric surgery because including all of these individuals and person-trials and then expanding them over the eight years of follow-up would result in a very large number of eligible person-trials to include in the analysis.

When doing this, we went from 770,000 eligible person-trials to 53,000 approximately person-trials due to this random sampling strategy. We did include all eligible person-trials among individuals who underwent bariatric surgery; and each of these analytical decisions and random samplings were accounted for in terms of the standard errors with our bootstrapping procedure.

And then taking these 53,000 person-trials who were assigned, who were resulted after the random sampling; as I mentioned, we assign each of those clones--one of those clones--to bariatric surgery within six months and another one of those clones to not undergo bariatric surgery within six months. For those who we assigned to undergo bariatric surgery, approximately 21,000 developed a contraindication and approximately 30,000 deviated from this their treatment assignment, meaning they did not undergo bariatric surgery within six months; those who developed a contraindication were kept in the analysis as would be done in a randomized trial, and these were not mutually exclusive because I wanted to demonstrate how many individuals develop a contraindication of how many people. There were ultimately then 24,660 person-trials who underwent bariatric surgery within six months or developed a contraindication, and this corresponded to nearly 13,000.

There were, again, clones who you then assigned to not undergo bariatric surgery within six months; again, the same number 21,000 developed a contraindication were kept in the analysis, and now a much smaller number of 4221 deviated from the assignment by, in fact, undergoing bariatric surgery. As such, 50,226 were included in this group and this corresponded to approximately 18,500 individuals.

This is a baseline characteristics table of the included participants in our study; the overall column is on the very left and then unweighted bariatric surgery and no bariatric surgery followed by weighted, as was described by bariatric surgery and no bariatric surgery that adjusts for baseline covariates as well as confounders that may be associated with both adherence and the outcome. These, I will reiterate, are covariates that were measured at baseline at time zero, whereas surgery could be performed or not performed within the grace period of six months, and these reflect individuals who, in fact, underwent surgery and just to give a better sense of the characteristics rather than those who also developed a contraindication would be included in that group as well.

Thing that I will point out comparing the second column of bariatric surgery with the third column of no bariatric surgery, both on weighted, I'll point out that the majority--a larger proportion who underwent bariatric surgery were female, those demographics with race and ethnicity were similar between bariatric surgery and no bariatric surgery; those who underwent bariatric surgery were slightly younger, more likely to have diabetes, more likely to have osteoarthritis and obstructive sleep apnea; the BMI was 4kg/m^2 higher approximately in the bariatric surgery cohort and they were also more likely to use oral hypoglycemics and insulin as expected. Among the diabetic subgroup, the hemoglobin A1c was 0.4 percentage points higher in those who underwent bariatric surgery compared to those who didn't.

After the weighting procedure in the two right-hand-most columns, you can see that these differences were largely resolved between the two groups with age being similar and there were still some notable differences in terms of bariatric surgery; in the bariatric surgery group, there were four percent more female individuals and but a lot of these differences were resolved and then we investigate these further in subgroup analyses. I'll also point out that among diabetics, there were 4 percent more in the bariatric surgery group compared to the no bariatric surgery group.

I'll present some of the results here, first in tabulated form and then in two figures. So, this endpoint is cardiovascular events at eight years. In the bariatric surgery group, there are 495 events and this is the number of unique individuals parenthetically; in the non-bariatric surgery group, there were 4700 events. These corresponded to, in the bariatric surgery grew 15.1 percent eight-year risk compared to the 15.6 percent eight-year risk in the no bariatric surgery group. These differences corresponded to an eight-year risk difference of 0.5 percent, being slightly lower in the bariatric surgery group and a hazard ratio of [0.4].

For the outcome of diabetes, the eight-year risk among individuals who underwent bariatric surgery within that six-month period was 16.4

compared to 33.1 among individuals who did not undergo bariatric surgery in six-month grace period; this corresponds to a risk difference over eight years of 16.8 percent favoring reduction in bariatric surgery, and a hazard ratio of 0.4.

I present these findings in a figure form here. In the left panel, the y-axis is the risk of cardiovascular events; for the left figure, Figure A, and risk of diabetes among non-diabetics for Figure B on the right-hand column. The dark blue line is bariatric surgery; the gold line is no bariatric surgery and months over eight years of follow-up is displayed on the x-axis.

Next, we present a subgroup analysis among diabetics only. Here we see a risk difference of 0.0 and a hazard ratio of 1.13; and this is graphically displayed here with the same structure of this figure as previous figures.

Next, we used, as another positive control similar to incident diabetes, change in A1c among diabetics. In this figure, this is a subgroup analysis among diabetics only given that they were the individuals most likely to have hemoglobin A1c measure. The change from baseline at the most recent value between three to five years was a decrease by 0.33 among those who underwent bariatric surgery, and an increase by 0.15 among those who were in the no bariatric surgery group. This corresponds to a mean difference in change from baseline at three to five years, of 0.48 or decreased among those in the bariatric surgery group compared to those in the no bariatric surgery group.

Next, as another positive control, we investigate a change in body mass index using a similar metric of change from baseline at three to five year, again, the most recent value in that time window. Those who were in the bariatric surgery group lost, on average, 3kg/m^2 of BMI, whereas in the no bariatric surgery group, they decreased their BMI by 1. This corresponded to a mean and difference in the change from baseline at three to five years of 4.4--that's a 4.4 or greater decrease with the bariatric surgery group compared to no bariatric surgery group.

I wanted to make sure that enough time was left for questions. And so, I will pause now and would be happy to hear your thoughts, comments, and questions. Thanks very much for everyone's time.

Heidi Schlueter: Great. Thank you so much. We are still waiting for questions to come in for the audience. I'm sure that there are questions out there; please use that Q&A panel in Webex, it's on the right-hand side of your screen, to type those questions in. We do have plenty of time for questions here, so we can get moving on those as soon as I start seeing any.

While we are waiting for those to come in, I know we do have Dr. Hernán on the call. Dr. Hernán, if you have any questions or any comments you'd like to make, you are able to unmute yourself and speak.

Miguel Hernán: Thank you. I don't have any specific questions here. I just would like to ask people to please ask questions because this is the first time that we're doing something like this with VA data and would be very interested in knowing the reaction. Thank you.

Heidi Schlueter: Questions are starting to come in; I'm going to start at the top and we'll work our way through. "Did you use the average of labs over the year for analysis?"

Arin Madenci: That is an excellent question. What we did was that we used--we did not use a running average--although that would be a very reasonable decision to make--we use time-varying laboratory values at each partitioned month of person-time over the first six months. We did allow _____ [00:38:57] period of time, but that would certainly be another parameterization if we felt that averaging the lab values cumulatively will be more reflective of the decision-making process of physicians who were deciding whether or not to offer a bariatric surgery.

Heidi Schlueter: Thank you. The next question here, "The lack of large differences in CV events at eight years seem smaller compared to what I recall from the broader literature. What are some thoughts about that possible difference?"

Arin Madenci: Thank you for that question. And that is what really intrigued me, intrigues us, keeps us motivated studying questions like this. I think there were a few important differences between the target trial that we published and the target trial that is implicitly emulated by previous observational studies. First, we allowed a grace period _____ [00:40:12] to undergo bariatic surgery. In previous observational analyses, these were equivalent to further propensities for analyses, target trials in which individuals were deemed eligible and assigned to undergo bariatric surgery on that day, which is not quite as realistic, one; and then additionally, it may not reflect the appropriate confounders when the decision was in fact made to undergo bariatric surgery.

Finally, we know that--and I did not show this here--but we do know that over the year before a bariatric surgery, there were improvements in creatinine, hemoglobin A1c, body mass index, and several other parameters that would not be reflected in the confounding control when it was measured at the time of the operation. And essentially, things, in terms of the analytical process and study design, are being done differently for those who undergo bariatric surgery and have a fixed time zero of their operative date and those who do not undergo bariatric

surgery in which they are often using the operative date of a counterpart in the bariatric surgery group as a reference, but is slightly more artificial in that way.

Heidi Schlueter: Great. Thank you. Next question here, "Can you discuss the differences with the prior propensity score analyses you presented in the background?"

Arin Madenci: Absolutely. So, that's largely what I--those questions are largely very similar. I'd like to mention another thing which is that the results that I presented here, that we felt were most realistic, and most reflective, and most mirroring of the previous randomized study that was, in fact, conducted among diabetics under one randomized trials with a six-month grace period, these results are not much different if we do not include a grace period. And so, it's not the grease period itself that lends to these differences because I think that's one added complexity--that is important, and I think most reflective of the target trial, but not necessarily driving the differences. I think, perhaps, the most important thing would be the eligibility criteria rather than just including all individuals within BMI criteria to have some laboratory and vital sign conditions under which they would be considered eligible similar to a trial.

And then this idea that the confounding pattern--the confounders--change over time because there is a time at which the decision to undergo bariatric surgery was taken; as I mentioned in one of the earlier slides, that's five or six months. And so, when that decision was made is truly at five months prior to the operation, and so I think there may be some residual confounding potentially at work there among other differences.

Heidi Schlueter: Great. Next question here, "Thank you, Arin. Great presentation. Could you speak a little more about your next steps?"

Arin Madenci: Absolutely. So, for our next steps, they would be twofold. One, we are really interested in expanding what right now are preliminary results due to, as I mentioned in the flow chart, two sampling steps that were performed as we familiarized ourselves with the VA system--the VA computing system--and figure out how to best optimize our computational efficiencies. And so, one next step will be to leverage the VA's capabilities on special servers in order to not have to sample and most closely emulate the target trial of interest. And so, that's what we are currently working on.

And then, as a second step, we are very interested in replicating the result of prior observational studies that were presented in the figures in the background section of these slides and in order to best understand

which deviations from the target trial led to these discrepancies in our findings.

Heidi Schlueter: Thank you. The next question here, "Just trying to see how best to use lab values for lipids et cetera."

Arin Madenci: Pardon me?

Heidi Schlueter: The question was, "Just trying to see how best to use lab values for lipids et cetera."

Arin Madenci: Okay. I will speak a little bit about the ways we use lab values and vital signs. Our goal was to--we essentially made the assumption that treating physicians who might refer patients to undergo bariatric surgery or to a weight loss clinic like the mood clinic, they may use lipid-based criteria or dyslipidemias in order to make those decisions; and so we also know that dyslipidemia is maybe associated with cardiovascular events, and so as such, this is both baseline and temporary confounder. And so, we tried to adjust for those lab values by weighting methods.

I'll mention that for all continuous covariates, we flexibly modeled them using restricted cubic splines in order to make a few assumptions about the relationship between those variables and adherence as possible. And if there are any specific questions other than that that I didn't answer, I'd be happy to answer them.

Heidi Schlueter: Thank you. Next question, "Can you review how you get from patients to patient trials?"

Arin Madenci: Absolutely, and this is a great question that I'll be happy to elaborate on. So, I think the best way to explain this would be to say that for all patients who were included--and again, I described that sampling pattern, but we have a number of patients who participants who underwent bariatric surgery and were included in that way and then a number of participants who were sampled and were known to have met eligibility criteria at least one point based on BMI level. With each of those unique individuals--with each of those veterans--we partition their follow-up time from the first time of eligibility until January 1, 2020.

We then, at each month during that time, evaluated whether or not they met eligibility criteria for our study based on what I described. And, for example, they could not have previously had a cardiovascular event, or died, or had a previous bariatric surgery; they must have lab values that were within the appropriate ranges, and they must qualify for bariatric surgery based on the BMI and comorbidity criteria that I outlined previously.

It is certainly possible for an individual to be eligible at one time point, not eligible at the next, and then eligible again in the future based on time-varying covariance such as laboratory values and those signs.

And so, for a given individual, they may have hundreds of eligible person-trials that they contribute and from which we sampled; same for an individual who underwent bariatric surgery, they would be eligible--potentially--during the six months prior to their bariatric surgery. And so, thinking of it, partitioned into months; and then for each of those months, indicating whether or not a given person-trial was eligible based on our criteria would be one way to think about this. And then, from that, each of those person-trials would be treated as a participant in our ultimate target trial emulation--and I described this in the way that we sampled 50,000 person-trials, and then included those with their full follow-up until they were transferred--or until follow-up was transferred.

I hope that answers your question. But if it does not, please ask again and I will try to clarify.

Heidi Schlueter: Great. Thank you. The next question here, "How do you write about causal inference study if you're submitting to a journal that prohibits causal language and observational studies?"

Arin Madenci: This is an interesting question. And I think that one of the benefits of describing the specification of the target randomized trial that we are interested in, and then stating that we did our best to emulate that target randomized trial, and clearly stating all the ways in which we kept the analyses similar and the ways in which we deviated from the analysis of our target trial as I presented in these slides here today, and then when we describe the effect estimates, those can be considered causal effects for the readership or journal editors and reviewers if they believe the assumptions that were made and believe that we were similar enough to our target trial within reason, or they can be considered to be non-causal, but we really refer to these--these are effect estimates of our target trial emulation and if the readership, in terms of published manuscripts, published papers, believes them to be causal, we leave that determination to stand on its own.

I would say maybe Professor Hernán may have more to add on that, but that is the approach that we took in this manuscript, in this analysis.

Miguel Hernán: I can add a couple of words to that if you want. I think that the fact that some journals don't allow you to use causal language is just a--is a misunderstanding. Because when you do an unobservational study, of course, you are not claiming that the final estimates are necessarily causal; they are our best attempt at estimating a causal effect. But there is no reason to to ban the use of causal language to describe what we are

trying to do; that's really where I think some journals need to change their policies is even if you are using observational studies, you are doing some type of analysis because the goal is causal and you have to be able to explain what the causal question is, in causal terms.

A separate issue is whether the final estimates have to be accepted as causal; and that, of course, is open to discussion as Arin said. But the fact that you cannot even explain that you are trying to estimate the causal effect is just a big misunderstanding.

Heidi Schlueter: Thank you. I received a clarifying question in here and somebody's just wanting to understand that, "You start with everyone who meets eligibility criteria for bariatric surgery then randomized to surgery or not surgery within an assigned time period that includes folks who develop a contra indication to surgery and remove those who are crossed over to the other group?"

Arin Madenci: Yes.

Heidi Schlueter: The next question here, "Were all measures collected from CPRS?"

Arin Madenci: I believe that they were, but that may be outside of the scope of my technical knowledge. The analysis was all done on VINCI using sort of the recently available data within the Clinical Data Warehouse. I hope that answers your question, but I suspect that it may not.

Heidi Schlueter: We will see if they follow up. Next question here, "Could you please explain the..." this is not a word so I'm not sure what it's supposed to be, "unstabilized adherence weights again and can you summarize your results? Were there significant differences in CV events at eight years?"

Arin Madenci: Certainly. So, in terms of the unstabilized weights that were used, the one thing to highlight, that I suspect is being asked about is that we simulated a target trial in which--and this was done using the numerator term during the grace period--we simulated a target trial in which for the bariatric surgery group, bariatric surgeries were performed in equal proportions over each of the six months of the grace period. The weights for the bariatric surgery arm, the denominator term--what I just mentioned had been the numerator term--the denominator term is composed of the probability of, in fact, undergoing a bariatric surgery given that one has not previously undergone a bariatric surgery given their covariates, both time-varying and based on coverage histories, and that they have not previously had the event.

In terms of the summary of the results, we found that the proportion of cardiovascular events at eight years was relatively similar in these preliminary analyses, around 15 percent. We did, in order to include

some positive controls, also investigate the proportion of diabetes--incident diabetes--among non-diabetics and that was very different at eight years, 15 or 20 percent greater among the non-bariatric surgery group compared to the bariatric surgery group.

And then, we had additional positive controls that we could compare to previous randomized trial data including hemoglobin A1c change from baseline, then change in _____ [00:56:50], and we saw essentially similar differences in change from baseline with bariatric surgery compared with no bariatric surgery to those previously published results.

We showed the diabetic subgroup, which had similar main results in terms of cardiovascular disease at eight years, and we are also planning to include other subgroup analyses as sensitivity analysis.

Heidi Schlueter: Thank you. The next question here, "Can you help me understand how the trial analysis handles patients that qualify for both the intervention and non-intervention parts of the target trial? I don't quite understand how you could have someone in both parts."

Arin Madenci: Absolutely. So, this is one of the analytical maneuvers that we must use because we do not, in fact, know as we would know in a trial--in a true randomized trial--what the treatment assignment is at baseline. And so, we generate a clone of each individual contextually and allow that clone to be part of both groups at baseline; so, one clone allows that participant to be part of both groups by having one clone in the non-intervention, as you mentioned, group and another clone in the intervention group. An alternative way to doing this would be to simply flip a coin and for each participant, if the coin lands heads, put them in the intervention at baseline; and if the coin lands tails, put them in the non-intervention arm.

I will point out that this is a less statistically-efficient procedure than to allow a clone to be placed in each group. And so, the reason that we include this complexity is for statistical efficiency.

Heidi Schlueter: Great. Thank you. We are at the top of the hour, so we are going to wrap things up. We do have a few pending questions here; I will get those sent over to Arin; but in regards to time, we are out of time for today.

I just want to open things back up to Arin, or Miguel, or Mihaela, to see if anyone has any closing comments you'd like to make before we close the session out today.

Arin Madenci: Thank you everyone for your time, and I'll be happy to answer any remaining questions by email or by chat. Thanks very much.

Heidi Schlueter: Thank you so much. We really do appreciate all the time that you put into your presentation today.