David Bekelman: Our next presenter Matthew Maciejewski.

Matthew: Just like it’s spelled.

[Laughter.]

Matt Maciejewski: Well, thanks very much. I’m going to try to go a little fast so I apologize for how quickly I’m about to speak but hopefully I can make up a little bit of time.

So this work was done in collaboration with a really great team that I’ve been lucky to have for almost all together six or seven years now. And this wouldn’t have been possible without HSR&D funding.

So this is the outline of where we’re going. The question that we wanted to answer, which is part of a larger IR is: do Veterans who had bariatric surgery between 2000 and 2011 in the VA have similar heath expenditures as Veterans who are eligible for surgery but didn’t have it? And we wanted to look at expenditures out up to five years if we could.

So the most common procedure that was done in the timeframe of this study is Roux-en-Y gastric bypass, which is in the red box. I’m a health economist not a surgeon so I’ll do the best I can here, the procedure essentially creates a smaller pouch in the stomach by essentially sewing the rest of it shut, which is the restriction part of the procedure. And then it’s sort of resecting the intestine, which is reducing the amount of absorption of food that can happen. So it’s a combination of a restrictive malabsorptive procedure.

The two other surgeries that were as common as this one is a lap band, a laparoscoping banding procedure that’s now completely fallen out of favor in the VA and outside. And a sleeve gastrectomy, which has now become the most dominant procedure in the VA and outside the VA.

So here is representing the two most commonly done procedures in the time period of the study. And you’ll notice that it’s open Roux-en-Y was the largest frequency of procedures. And open procedures are essentially not done any more, it’s all laparoscopic at this point. So there is some historical artifact in some sense to what I’ll present.

So the background of this cohort was we evaluated some other things in terms of mortality showing that surgical patients has longer survival than matched non-controls. And also it has much greater weight loss up to 10 years, 11 years, compared to matched controls here in the light blue bar.

And so the question that we were really wondering is: do the health benefits, the tremendous health benefits, of bariatric surgery translate into lower health expenditure? You would think that they would, it seems fairly logical. And there is some evidence in non-VA studies earlier in the timeframe of this limited literature suggesting that there are lower expenditures among surgical patients compared with non-surgical patients out to three to five years. But you’ll note that most of these studies are done on mostly women. This is a minority of the males, in their forties and some in the early thirties as well.

There are a number of other studies as well, including one that we did. And then an earlier \_\_\_\_\_ [00:03:08] (IER?) showing that costs are actually similar, that’s the key here. And this is a Blue Cross Blue Shield population, this is Veterans, this is also Blue Cross cohort.

So we conducted this follow-up study from this earlier one using a retrospective cohort study as we had previously, with contemporaneous but non-equivalent controls. We identified the surgical patients from … Originally we had used the VASQIP data but in looking at VA data directly, back in the days of Austin, we found that we identified surgical patients with CPT codes and ICD-9 codes whereas VASQIP only identifies patients with CPT codes. And so there were some cases that we captured. We also got surgical patients from fee basis, which VASQIP does not.

So we kind of did our own home-grown cohort identification. And we identified controls in terms of Veterans who are eligible for surgery on the basis of BMI because we have all this great grey data. We started out with a sampling frame of almost 1.45 million Veterans over this timeframe. This is a graphical illustration of these cohorts by year.

And one thing that I want to note here is that a non-surgical control could arise in any or all of these different time periods. And so when we’re doing matching, we want to allow for this possibility for sampling with replacement. And conventional methods for matching don’t really handle that very well. So that was something we wanted to think about. Also, some non-surgical controls crossed over into the circle cohort, we had to handle that as well.

So there are a number of inclusion and exclusion criteria that I’m going to, for the same of time, not go through, but trust me they’re all reasonable.

[Laughter]

So, then we had this problem with matching over a long cohort of patients. If we just had a single year, say it was just a single year, we could do conventional propensity score matching defining everyone, and just defining people on a given single limited narrow timeframe. But since we had these multiple cohorts, if you will, we then considered alternatives.

Calendar time-specific propensity score matching is a relatively new method and extension of conventional propensity scores that was developed by Mack and Brookhart and colleagues at UNC. But we didn’t do that for reasons I can explain if you come talk to me afterwards.

There was a newer method that my colleague Maren Olsen identified called sequential stratification that is specifically designed for handling this kind of complex, longitudinal matching. And so that’s what we did.

It has several advantages. It allows for possible multiple index dates, say for these controls that arise over time, because they have many, many measurements over many years that make them eligible. Because otherwise, if each point at which they have an eligible BMI, if we had to do something like, well what’s the best one? That can be an intractable problem. There may be a lot of best ones. It’s just that the best one may match the different patients over time, both in time and on the characteristics around them, and that’s the basic idea of this approach. And we can match over time as I sort of just said. And then this method also handles crossover when controls become surgical patients by censoring them essentially.

There are some limitations to this. The main one is that it’s not a \_\_\_\_\_ [00:06:49] program so we had to do this kind of manually. And also that covariate matching, unlike conventional approaches to matching where you just throw everything in the kitchen sink into your treatment allocation model, you have to in this approach sort of identify apriority what you think are the key confounders that are going to be important in sorting patients into surgery or not.

So the basic set up of this is that, think of this as a series of N of 1 trials where we’ve got each surgical patient ordered from the earliest surgery to the most recent surgery. And then we sort all the patients who are controls by their weight measurement, the date of their weight measurements. And then we simply sort them in time.

Then you can think of a range of co-variates, and we want to identify for each surgical patient which group of controls who have eligible weight measurements close in time look like them, is the basic idea. This is out of an idea by Paul Rosenbaum, called Risk Set Matching, if you’re interested in the original source of this idea.

And so the specific things which we identified the patients as being like the surgical patients were: gender; diabetes diagnosis, we could do exact matching on all these binary variables; race via region, because we wanted to have co-located this as best we could by VISN; and then BMI categories, as a first pass to cull this large potential match cohort down, BMI categories in groups and then age within five years, buckets.

And then to further restrict, because there ended up being anywhere between nine potential controls up to 222,000 for any given surgical patient, we wanted to narrow that down to those who really looked very much like the surgical patients.

So then we applied a Mahalanobis Distance Function to the three closest matches on continuous covariates, which were co-morbidity as measured by the DCG score that was very handily available, age, and BMI. And so with that we got up to three matches for each surgical patient.

So the expenditure outcomes that we’d examine were the probability of an in-patient admission. We could not estimate in-patient expenditures in a two-part kind of set up because the costs were just so scarce and the distribution got really sparse out of the right tail which essentially made it impossible despite many, many months of attempts. So we were limited to looking at the probability of admission.

We looked at out-patient expenditures, out-patient medication expenditures, and then total expenditures, rolling everything up. So in-patient expenditures are represented in total expenditures but we don’t show it separately. And we inflation adjust it.

Then we examine these outcomes in the surgical patients and the controls in six-month buckets. And that was chosen to essentially avoid having costs that were diced up so finely that everything had a high proportion of zeros which would have forced us into the two-part model world.

We looked at cost three years prior to surgery and five years after, which is the same basic set-up as we did in the prior work. I should note that we excluded all costs, especially in the total cost analysis in the six months leading up to surgery because the cost of the surgical admission was in that bucket, and it was such a humungous spike we just couldn’t model it. So that’s problematic in some sense.

So we did a basic descriptive comparison, then we estimated a bunch of fancy models, let me just put it that way. We used two-part models, I should note, for the pharmacy and outpatient expenditures using marginalized two-part method that other great statistician on the study, Valerie Smith, developed in her dissertation. And then we did difference in difference.

This is a summary of the patient characteristics. Two things I want to point out, it’s 75 per cent men, I don’t show that gender thing for some reason, but the mean age is 52, 53. And the other thing is that covariates are well balanced. Standardized difference is less than 10 and absolute value indicate good balance. So we balanced well in the things that we included.

We also inadvertently, or fortunately, balanced on things that weren’t in the model in our selection criteria, probably because they were highly correlated with things that were. But then there were a host of other things that were not well balanced. We didn’t include them in our treatment allocation matching and so we had to adjust for those.

What I’m showing you here, at the top is the estimated or predicted expenditure of probability of admission for the surgical patients in light blue and the matched controls in the salmon color. And below we show the difference in difference estimates in each six-month block. So the bottom line here is that in the baseline in the pre-surgical period, probability of admission, was the same as we would hope it would be if we were pretty well matched. And probability of admission diverged because it was higher for surgical patients in the first few years after surgery and then converged to be the same afterwards.

For outpatient expenditures, they were surprisingly different in the pre-surgical period, the surgical patients having accelerating out-patient, increasing outpatient expenditures, and then that remained stable in the post-surgical period. This was a surprise to us. But when you put this in a difference in difference context, there is really no difference in difference. But this result of stable consistent out-patient expenditures differed from all the prior literature out there, including our earlier work.

But one result that was consistent with the prior VA literature and non-VA literature is that medication expenditures for surgical patients drop like a stone, presumably because their diabetes and hypertension and other conditions improved and maybe they could go off drugs almost entirely. And so when we look at the difference in difference, it’s highly significant.

But the one thing I should note though is that the dollar amounts here are fairly low. These are all out-patient expenditures. So the things that they are coming off are essentially cheap drugs, so the return on investment, even though the effects size is huge, in dollar terms is modest.

And then with total expenditures I just want to show you the unadjusted total expenditure. This is the humungous spike due to the surgery itself. And this shows the unadjusted trend in total expenditures for the surgical patients and then for the controls.

Am I saving us some time here? Not really.

And so then for total expenditures, they are fairly comparable in the baseline period and then converge to be similar essentially two years out, which is consistent with our prior work but inconsistent with some of the earlier non-VA work. So I think the true believers would expect this blue line to drop like a stone below the salmon line for the controls for medication expenditures. But for total cost it just doesn’t do that. And we have lots of ideas for why that may be and hope to be able to examine that in future work.

So there are a number of limitations. This isn’t generalizable beyond 2011, probably not outside of the VA where a predominantly older male cohort were evaluated. Open Roux-en-Y is the dominant procedure as I showed you, so it’s somewhat of a historical artifact. And there’s some research confounding running through this even though we tried to do a good job matching, as I showed you there are some imbalances in some of the observed factors so presumably there will remain some unobserved things.

In conclusion, bariatric surgery was not associated with lower expenditures out to five years, except for medications, but the dollar amount is modest. So in terms of next steps, we’ve actually decided to take this out to 10 years because we published this weight change paper last year out to 10 years and it seems only logical to follow it up. So hopefully in a couple of months this will all be updated and we’ll be able to see whether that blue line crosses the salmon line out to 10 years.

And we’ve got new funding to follow the same cohort looking at opioid misuse, alcohol misuse, and depression treatment. So that’s it. Thank you.