MCH EPI Conference

Using Population Attributable Fractions to Assess MCH Population Outcomes

December 8 – 11, 2008

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: I never used the beta for poverty, right. Because right, I wasn't calculating the relative risk for poverty because we are thinking of it as unmodifiable. All it is doing here is, poverty is doing here is adjusting the coefficients for smoking, cocaine, and the interaction term. The only reason we end up using the interaction term is because smoking is involved in the interaction term. Wanda…

WANDA: (Inaudible)…

KRISTIN RANKIN: The summary page…I'm sorry the proportion of low birth weight that can be reduced is by eliminating smoking and cocaine in the population after controlling for poverty is sixteen percent, I'm sorry fifteen point five eight percent.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: I'm sorry? Which one? Oh, all I've calculated here was for sequential PAF's. I haven't gotten to the average yet. This is the first sequential PAF for smoking and this is the first sequential PAF for cocaine. Actually I could just go ahead and do that right now because I think that is a good question and I think I'm…oh, it's right here. This is the first sequential for smoke. Then for cocaine, the first sequential is point one five five eight. In order to get the…when smoking is eliminated second, you would subtract the summary PAF, which was point one five five eight, I'm sorry this is the summary PAF. In order to get…this is the risks systems, smoke is a variable of interest. This is smoke removed first and that is the first sequential PAF, which is the same as the adjusted, which is point zero seven three five. Then this is smoke removed after coke has already been removed, smoking is removed second. So in order to calculate smoking removed second, you take the summary PAF for smoking and cocaine, subtract out the first sequential PAF for cocaine to get what is left over for smoking. Then you would average the first sequential PAF for smoke and the second sequential for smoke to get point zero six five seven. So it is six point five seven percent is the average PAF for smoking. I want to show you this set up, I want to go back and explain what I've done here because we are going to use this throughout and throughout the modeling exercise to. We wanted to really lay everything out because this gets more complicated as you move beyond two factors. So you have these are the risk systems of smoke is a variable of interest, the removal of ordering. You know that with two variables in the model that you just have two ways that you can remove it, first or second. We have the summary PAF for each of these, and the summary PAF when

each of the variables interests is not part of the risk system. What that means is, in this case, this is the summary PAF for smoking and cocaine, since cocaine was removed before it. This is the summary PAF for when smoking is not part of the system. This is when cocaine is removed first. Then you get your sequential PAF's when you subtract this from this. Then there is only one sequence that shares the sequential PAF and I will show you why that is important later. Then you do the average over them. I'm sorry; you multiply the sequential PAF by the number of sequences, which in this case it is just multiplied by one. Then take the simple average of the number of sequences and the sum of the sequential PAF. So for cocaine we do the same thing, if it is removed first it is point zero nine eight zero, which we just calculated in the previous slide. If it removed second, you calculate that by taking the summary PAF for smoking and cocaine use and subtracting the first sequential PAF for smoking to get point zero eight two three. Then you are going to average the first sequential PAF with the second sequential PAF to get the point zero nine zero two. In this case it is pretty simple, it is just two variables and it is a simple average of those two sequences. Are there any questions about this before we move onto three variables?

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Oh, is it? This is a good breaking point. Ok, thanks.

DEBORAH ROSENBERG: A couple of summary points before Kristin goes into the next scenario where poverty we'll consider modifiable. One thing is this term we are using "risk system," really the factors is the risk system are the modifiable risk factors. That is why in the smoking, cocaine, in the simple version that I talked about, not controlling for anything else, had two sequences, and two sequential PAF's within each. That didn't really change when we are considering poverty as unmodifiable because poverty isn't really…we haven't yet added another variable to the risk system, per say. That is what we will move into as Kristin…as we consider poverty modifiable, and then we will have a three factor risk system, that is why the number of sequences and the number of sequential PAF's and all of that will go up. But so far we really have just been in two variable risk system and when you look at this piece of the Excel file that Kristin was talking about in the end, creating the averages and how you do the intermediate calculations, you start thinking about when variables are the variables being controlled for, even the modifiable ones temporarily they are kind of not in the risk system, so variables are moving in and out of this risk system depending on the combinations being controlled for and the combinations that are being adjusted for simultaneously, so that is why it gets more and more complicated. In this example we would have when it is smoking controlling for both cocaine and poverty, really only smoking is in the risk system. When it is smoking and cocaine together in the risk system controlling for poverty and there never was smoking, cocaine and poverty because we haven't put that in the risk system, yet, but that is what is going to happen next. So we will do the next scenario, be able to compare the results to what we just did.

KRISTIN RANKIN: So let's just go back to the PowerPoint slide for just a second and summarize what we just did. So we found that with smoking and cocaine, this is sequence one. You have a sequential PAF for smoking of point zero seven four, that is point zero seven three five. A PAF for cocaine use of point zero eight two, that was subtracting the first sequential PAF for smoking from the summary PAF for smoking and cocaine. Then you also have sequence two which cocaine then smoking, where the sequential PAF, the first sequential PAF for cocaine was point zero nine eight. The second sequential PAF for smoking was the summary minus the first sequential PAF for cocaine, which was point zero five eight. So in the next slide you will see how these become part of the pie. This is sequence one and this is sequence two. They are both controlling for poverty. So you can see how theses sequential PAF's are different, its point zero seven four, the first sequential PAF for smoking is point zero seven four and the second sequential PAF for smoking is point zero five eight. So that is why we want to take an average of these. This is just a reminder of what the average PAF is, it is based on game theory for risk allocation and here is a generalized equation for it. The one over N factorial here, N is the number of factors that you have, so in this case it is two. We have two factors in the risk system. So this is just one over two. It is just a simple average of the two sequential PAF's, but because this is a factorial, it will get more complicated when you have more risk factors, because you'll have six sequences and there are three factors, you will have twenty four sequences when there are four, and so on. So you can go back to this to just generalize. But here's the calculation of the average PAF for smoking. The first sequential PAF, the second sequential PAF for smoking, and the average is point zero six six. We showed this in the Excel file E, Excel

file F, maybe. Then the average PAF for cocaine is point zero nine zero. I don‘t know what happened to the pie charts. We used to have pie charts for that in there. We compare them all later, so we will see them all next to each other all later.

So let’s go on right now to the sequential and average PAF if we considered poverty modifiable. So now we have three factors in our risk system. As I mentioned that there are six equations. I mean six sequences then. Smoking removed first, before cocaine, then poverty. Smoking removed first before poverty, then cocaine. You can see where this is going to be the first sequential PAF. The value of this is going to be the first sequential PAF for sequence one and for sequence two for smoking, but we have to count them in twice because they are two unique sequences. Even when I just said from my dissertation that there we seven hundred and twenty sequences, there are only thirty two unique values for sequential PAF’s, it is just that you have to know how many times that they come up. So how many time is smoking removed first, if this was a six factor risk system, remember we are looking at smoking removed first. It is going to removed first how many different times, twenty four different times; it could be removed first, depending on the organization of the other variables after it. So that is an important concept to remember for later. For the third sequence, smoking removed second after cocaine, but before poverty, it is going to be different than smoking removed second in sequence five after poverty and before cocaine. These values will be different, the sequential PAF’s for smoking would be different. When smoking is removed last it is in sequence four and six. Those values would be the same, but we could for them twice because there are two unique sequences there. Any questions about that? Ok, but it is

just important to know so when there are three modifiable factors in the risk system there are six sequences. You are going to take an average over the six sequential PAF's for each one of those factors. It is the same model though. We are using the same model to estimate this, but now we are going to use the beta coefficients in a different way. It just shows the factorials, I jumped ahead of myself here. If there are four factors there would be twenty four unique sequences, five factors, one hundred and twenty sequences.

Then to calculate the sequential PAF's for factors removed second and third in a three variable risk system it is necessary to compute the PAF for every pair of two factors, adjusting for the third factor. We are calling these intermediate summary PAF's. We struggle with the terminology for this because there is not really a word in the literature for this. I don't think (inaudible) never really names this, but the intermediate summary PAF's are necessary in order to calculate all of the sequential PAF's and you will see why.

We are going to Worksheets F and G in the Supplementary Excel file. So here we are calculating the first sequential PAF for smoking, cocaine, and poverty, where poverty is modifiable. So it is a three factor risk system. These are the results of the binomial regression, the models, the beta coefficients are exactly the same as last time, but we are going to be using them a little bit differently. You see that the adjusted relative risk did not change for smoking, the adjusted relative risk did not change for cocaine, but

now we have calculated the stratum specific relative risk for poverty. Where in the strata of smoking is yes and smoking is no. So you can see that it is one point one three five one where smoking is yes. And two point one seven nine where smoking is no because of that interaction term in the model, the smoking and cocaine interaction term. So now that I have the adjusted relative risk for cocaine and poverty, I have the PJ, which hasn't changed through all of these examples. I can calculated using the Bruisey method, the PJ over RRJ. So these are the different pieces that we need to calculate the summary PAF and the different average PAF. Down here in the second row here, I did this so you could print it out later. We also need to calculate the relative risk for every combination of these two factors. So what is the relative risk for smoking and/or cocaine controlling for poverty? What is the relative risk for smoking and poverty, and/or poverty, controlling for cocaine? So we need these relative risks, think of it as when Deb was explaining earlier, the four by two table, but we are going to be using all of these, all of these relative risks to calculate the intermediate summary PAF's, which we are going to need for the average and sequential information. To get the relative risk for smoking and cocaine, we want to apply the beta coefficient to the code for smoking and cocaine controlling for poverty. So here you can see that you get several different values for the relative risks and this is again because of the interaction term in the model. Your referent group is going to be where smoking and cocaine is zero. So you can see that is why the relative risk is one in these last two rows because that is where smoking and cocaine is zero. That is your common referent group. You get a four point two zero nine, where just cocaine only is the exposure. You get…this is really complication with this interaction term. You get point seven seven six, where smoking is one and cocaine is

zero. One point one zero where smoking is one and cocaine is zero. So you see that one point one zero comes up again because of the adjusted relative risk for smoking, controlling for cocaine and poverty. In the second row here, what you are getting is the joint affect of cocaine and the relative risk for that is seven point seven four three. The first one is also the joint affect, but it is in the stratum of poverty equals one. So you are getting a lot of different relative risks. Normally you would just get the three, for smoking and cocaine, smoking alone, and cocaine alone, but because of the interaction term, you also have the smoking alone where poverty is yes, smoking alone where poverty will be no , and then where smoking and cocaine within poverty equals yes. Then smoking and cocaine within poverty equals no. Is it making sense how we got those relative risks over there? Any questions? It becomes more…I guess these are all very complicated by those interactions. Since smoking…which one should be the easiest? Yeah, it is only cocaine alone because that is the only one that doesn't have an interaction term. I should have done and example without interaction, it would have been a more simple, but now you have the most difficult example so when you come across an interaction in your data you can use this. Basically for smoking and poverty here, actually this ends up being the most simple example where you only have three different relative risks. In this case you have smoking and poverty in the stratum of cocaine equals one. You have smoking and poverty, no poverty. You have smoking and poverty and cocaine equals zero. Smoking in the stratum where cocaine and poverty are both zero. So once you calculate all these different relative risks for the different combinations of factors and you have all three factors at the end. Which these relative risks should look familiar because these are all the common references of no exposure.

Then you can calculate the PJ over RRJ for each of these combinations. You are still applying the proportions of the cases in that stratum times the relative risks, I'm sorry divided by the relative risks. So this is for smoking and cocaine, you add across all of those and subtract from one and you get the intermediate summary PAF for smoking and cocaine controlling for poverty. So you've seen this before, this is the same as the summary PAF that we did when poverty was unmodifiable, it's point one five six. The summary PAF for smoking and cocaine controlling for poverty, but then we also have the summary PAF for smoking and poverty controlling for cocaine and the summary PAF for cocaine and poverty controlling for smoking. We need to get all of these in order to calculate the sequential PAF. Then the summary PAF is a lot larger now it is point four four one, when we are considering poverty as modifiable. That goes back to the fact of the fake dataset that we have so many people that are in poverty, the presence of poverty is very high. The relative risk is high as well. So our summary PAF jumps all the way up to point four four one. We have all of these other building blocks that we are going to use to calculate sequential PAF's. I will show you how to calculate the sequential PAF and then the average PAF's. I will show you how to do that in the next one. If you have questions about those relative risk calculations you can…we can talk about those if you want to take a closer look at those at the break or before tomorrow we can talk about those. They are complicated by that interaction term. Maybe what I will do tonight, I will make an example without an interaction term, we can walk through that and it could be a little bit more simple.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Uh huh. You can see even in the first sequential PAF for poverty it is point two seven five, which is a lot higher, twenty seven point fiver percent, which is a lot higher than the seven or nine percent that we've been dealing with in smoking and cocaine. Ok.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: So we are thinking of a very high risk subpopulation of the…whatever your jurisdiction is. Maybe you are looking among…it depends on eligibility levels, but maybe you are looking at Medicaid women or some higher risk population. It is the poverty line here that is your eligibility for Medicaid. So then you can see how this becomes a little more complicated when you are doing the average PAF.

So now we have several risk systems with smoke as a variable of interest. We have smoke removed first, and as I mentioned there is two different ways that smoke can be removed first before cocaine then poverty or before poverty then cocaine. So I've taken account of that in this column where they are sharing the same sequential PAF. This makes your Excel spreadsheet a lot smaller rather than doing every possible sequence, especially when you get to four and five variables. This will help us especially later, but there are two sequences that share the same sequential PAF of removing it first. Then

you can remove smoking second after cocaine or second after poverty and you can remove smoke third after cocaine and poverty in either order. So again this sequential PAF would happen twice. I’ve accounted for that here and I will show you how that factors into the equation. Now we have this intermediate summary PAF for the variable of interest, smoking and any factors removed before it. So this is the first sequential PAF for smoking because no factors were removed before it. This is the sequential PAF for smoking and cocaine. This is an intermediate summary PAF. This is the intermediate summary PAF for smoking and poverty, which we calculated on the last spreadsheet, point three eight three. Then the summary PAF of the entire risk system, for all three factors is point four four one. Then you want to subtract here the summary PAF when the variable of interest is not part of the risk system. You subtract zero because from the first sequential PAF there is nothing in the risk system if you remove smoke. In this case you would subtract from the intermediate summary PAF for cocaine and smoking. You are going to subtract the first sequential PAF for cocaine and what is left over is the sequential PAF for smoking removed second after cocaine. For the smoking removed after poverty you are going to take the intermediate summary PAF for smoking and poverty and subtract out the first sequential PAF for poverty alone, which is point two seven five, and you will get the second sequential PAF for smoking after poverty is removed first. Then you take your summary PAF is point four four one for all three factors then you want to subtract out the intermediate summary PAF for cocaine and poverty to get what is left over for smoking. This is why we needed all these different pieces in order to calculate the sequential. Basically, what you are doing is trying to see what proportion of the pie, what proportion of the summary PAF of all the factors is left

over after you've removed each step. You get the sequential PAF in this column here. You multiply them by the number of times that they come up, which is as we said the first sequential PAF comes up twice. I've done that here, this is the product sequential PAF in the number sequence. Then in my final equation here, I've summed over all of these values and divided over the number of sequences, which is actually six, the sum of this column. So I've divided the sum of this by the sum of the number of sequences, which is six. I've gotten a smoking average PAF of point zero eight zero nine. So then you go back and do the same for cocaine. Cocaine removed first then second, then third. You are using the same set of values, but in different places. You will see how they may come up. The first sequential PAF for cocaine, then you are using smoking and cocaine, poverty and cocaine, then the summary PAF again for all three factors. You will have to calculate all the possible sequences for cocaine and then average over them, the average PAF for cocaine. Then you will do the same for poverty, the average PAF for poverty ends up being point two eight one. I have a summary of that in the slide.

You have sequence one is cocaine and then poverty. You have the values here. So the yellow is always going to be the smoking value. The blue is the cocaine and the purple is the poverty. We have sequence one and sequence two. These are the high parts of sequence one and sequence two. This is illustrating why the first sequential PAF for smoking is the same for these two sequences, but you can see how cocaine is different when it is removed second, than when it is removed third after smoking and poverty. Now the total colored area is not changing, the summary PAF is always staying the

same. It is how you are distributing that is changing all these sequences. Then you have sequence three and sequence four. You can see how the different variables change here. Where cocaine is removed first there is a point zero nine eight. Then there is a small slice for smoking here of point zero five eight. Then whatever is left over of the summary PAF is attributed to poverty. Because we have these six different pie charts, that is why we want to go on and calculate an average PAF and get an average of all these sequences for the purple, which is going to be an average for poverty, an average for the yellow, which will be the average for smoking, and an average for the blue, which will be for cocaine. This shows how I calculated it, in the Excel file. It is the average PAF and I have sequence 1A, sequence 2A, sequence 3B, and you will see that those correspond if you go back in the slide, all the different sequences that I described in the beginning. So you are just rearranging at this point, all the sequential PAF's, all the sequential for smoking, and you take the average of those six. You have all the ones for cocaine; take the average of those six, and so on. Go ahead…

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Uh huh. Here, the two one one two? Because smoking can be removed from the population before cocaine, then poverty, but it could also be removed first from the population before poverty, then cocaine, so even though it is the same value for the first sequential PAF, there are two separate sequences. We have to take into account that there are two different sequences that could be removed.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: I'm sorry, go ahead.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Right, because you have the sequence for smoking, cocaine, poverty, and smoking, poverty, cocaine. So you have to count that twice, and then it could be removed third twice, cocaine, poverty, smoking, then poverty, smoking, cocaine. So you have to account for each of those even though it is the same value.

UNKNOWN SPEAKER: (inaudible)…

KRISTIN RANKIN: Sure. Ok. Yes, these are the six possible sequences for a three factor risk system. See how you have smoking. Sequence one and sequence two are both going to have a first sequential PAF for smoking. Then sequence three has smoking removed second and sequence five has smoking removed second. The value of this PAF is slightly going to be different since this smoking sequential PAF is going to

be a result of smoking plus cocaine minus cocaine. So what is left over from the intermediate summary PAF of smoking and cocaine considered together, what is left for smoking after cocaine has already been removed? Rather than down here you are going to have the intermediate summary PAF for poverty, smoking, which is much larger because poverty has such a large PAF. What is left over for smoking after poverty is removed is going to be a different value than three. So that is where the second sequential PAF will have unique value that is why I only multiplied that by one. Then when you go to the third sequential PAF, basically in sequence four you are going to have the summary PAF for the entire risk system for all three factors minus the intermediate summary PAF for cocaine and poverty equals the third sequential PAF for smoking. In this case you have the summary PAF for all three factors minus the intermediate summary PAF for cocaine and poverty, which is the same value as right here. Then you are going to get the third sequential PAF for smoking. That is why that will play in twice. Other questions, go ahead.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Yes. Uh huh. Right. Ok, so for smoking, it would be eight point one percent of low birth weight could be reduced in the population if smoking were eliminated. I'm sorry, counting for cocaine and poverty. So on average eight point one percent, I think it helps if you say on average; eight point one percent of low birth weight

could be reduced by eliminating smoking in the population after controlling and considering cocaine and poverty.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Yes, and the same for all the others. So what this is allowing you to do is that we are kind of working backwards from the summary PAF. We want to make sure that we are not getting…if you added the first sequential PAF for smoking to the first sequential PAF for cocaine to the first sequential PAF for poverty, you would get a much larger value than the summary PAF of point four four one. We don't ever want to go over that value because we know that if you eliminated all three simultaneously, that all you can do is eliminate forty four percent of low birth weight. So you don't want to add those first sequential PAF's, we know that those are going to be an overestimate. So you start from the summary PAF and kind of work your way backwards and say how do we apportion this summary PAF and the average is the way that we come up with to do most fairly so that you can then compare smoking to cocaine and to poverty and say "In this case if you eliminate poverty you get the biggest bang for your buck," right. It is about twenty eight percent of low birth weight could be eliminated by eliminating poverty in this scenario. So you might think about risk factor reduction strategies by prioritizing these or ordering these. Now a lot of other things are going to go into that decision making but this provides a way to quantify the…

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Uh huh. Right. It did. Deb and I were talking about that. Do you want to provide an explanation?

DEBORAH ROSENBERG: Sure. Well I don't know if it is an explanation, but actually first I thought it was kind of strange to, but…So if you look at…it is interesting that when we are not considering poverty as a modifiable factor, let's just focus on smoking here. So we have a PAF for smoking of point zero seven, when it was just smoking and cocaine not even controlling for poverty. Then this is controlling for poverty, but poverty is being considered unmodifiable, and smoking is point zero six six. That kind of intuitively makes sense that we are doing adjustment and diminishing what the PAF is. Her we are considering poverty modifiable, it is taking up its share of the pie, and when you look at the smoking slice, it has actually gone up now, and it is eight percent. We were wondering about this and talking about this in relation to an article also and here is what I am thinking, I don't know this, this is something that we are thinking about, but we have a lot of unexplained, unattributed outcome here. If we could imagine that if we ever, which we never will, knew every causal mechanism for low birth weight so that there wouldn't be any blank part of the pie, it might really change. So we don't know in these pies, there might be other causal mechanisms out here that include somehow being related to smoking that we don't know about and that we are not accounting for that somehow as we enrich our model and consider more things modifiable, again we

are thinking even considering poverty modifiable, we know that it really is a marker, who know how many actual modifiable risk factors. By doing that, considering it and disentangling it from smoking we've actually ended up with a higher, bigger pie slice for smoking. The problem is that we are never going to be able to complete any of these pies, no matter what we do. So we could add ethnicity if we wanted to, consider it modifiable, as a marker for truly modifiable factors. We could add everything, hypertension, all the know risk factors for low birth weight. Actually we will do more of this tomorrow when we do variable modeling. We are never going to have everything. We don't know fully all the causal mechanisms. That is going to mean that we have some imprecision. It's like any other thing that is why we need ninety five percent confidence variables. We have imprecision in these estimates and we, right now, don't know what the variances are.

KRISTIN RANKIN: I'm wondering if the impact of the smoking and cocaine interaction as well, is kind of impacting it.

DEBORAH ROSENBERG: You mean smoking, poverty interaction.

KRISTIN RANKIN: I'm sorry smoking, poverty interaction.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Uh huh.

DEBORAH ROSENBERG: Right. Sure because then we are using stratum specific…see that is another interesting point ,actually, thinking about interaction in the context of the PAF. Although, we never want to summarize over heterogeneous, odds ratios are relative risks. Think about heterogeneous stratum specific PAF's. So if you have a PAF in one stratum that is…I'm making this up…let's say three percent, and a PAF in another stratum that is twelve percent, but for the PAF it is still going to be appropriate to take some average of that because you are actually talking about the burden in the population. Even if you had in one stratum, the extreme example if something was a protective factor in one stratum and a risk factor in another stratum, so you might get a PAF that is less than zero, but what does that mean, you still want to know, on average how much are you going to reduce the outcome that you are interested in if you eliminate the risk factor in the population. So it is a little trickier. The interaction is a little trickier in the context of PAF than when we are just thinking about the ratio measures.

KRISTIN RANKIN: If you are thinking about as you add factors to the pie, you are not going to get easily up to one hundred percent because those factors are going to overlap with these other factors in the population because people in general, the population, is going to have a number of different people with different risk factor

profiles. You could be smoker and a cocaine user and you have several other risk factors, as you add to those the smoking is actually going to get reduced if you add a factor that overlaps quite a bit with that. That will take a proportion of the smoking PAF. It is not that when you add factors, in this case when we added poverty, we came up with a lot more explained, but this is an extreme example. Usually as you add factors your summary PAF isn't going to get a great deal larger. So you are not going to be filling this pie up very easily.

DEBORAH ROSENBERG: Also. I think here we have identified what is going on here. It has to do with the interaction term in the model.

KRISTIN RANKIN: Not the magnitude of the poverty though.

DEBORAH ROSENBERG: No, the magnitude of smoking.

KRSTIN: Exactly.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: Yeah. Well we just talked about a little bit, but yes you might want to…here again we have left a product term in the model, but just like in regular modeling you will make a decision of do you want to run stratified models. We only talked about it for a minute when we referred to the work that Kristin did for her doctoral work, where looking at obesity in the National Survey of Children's Health. She ended up stratifying by both race and gender, rather than putting…because she got a lot of significant interactions. It was going to get too complicated, not only complicated mathematically, but not iterpretable. So it was much more reasonable to run stratified models.

KRISTIN RANKIN: I think that it really depends on, again your conceptual model and what you want to get out of the analysis. In the case of my dissertation, I was considering race ethnicity as an unmodifiable factor; it interacted with several of my modifiable factors. Sports participation for girls and boys had different impact on obesity, as you would expect, or physical activity as well. When I saw interaction between unmodifiable and modifiable, I wanted to stratify by the unmodifiables to describe it in those different groups.

In this case if you are considering poverty modifiable, it would be an interaction between two modifiable factors. You just want to decide what your interpretation is going to be. Do you want to describe the PAF for cocaine and poverty amongst smokers and among

nonsmokers? Is it more meaningful top get an average PAF across the population, but you want to use the interaction term in the model because it is the best model.

DEBORAH ROSENBERG: And again, that is actually a good example because just similar to the regular modeling we do, any time that you decide to stratify a facto, in this case, you are going to stratify a modifiable factor, you are never going to be able to report the impact of that, you are not going to have a PAF for that factor anymore because it is defining the strata. Stratifying by race, ethnicity, considering it unmodifiable, that is ok because you weren't going to calculate a PAF in that case anyway. Same thing here, we could have stratified on poverty, if we are considering it unmodifiable, and it wouldn't really change anything because we were deciding not to report a PAF for poverty. The minute that we decide that it is modifiable, and then it makes that decision about stratification, you have to think it through and decide how you want to report things.

KRISTIN RANKIN: Yes.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Uh huh.

DEBORAH ROSENBERG: Right.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: Hypertensive disorders.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: Preeclampsia. Not poverty, forget poverty. He is saying substitute for poverty, a real causal.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: Well you are actually bringing up another issue. When Kristin and I did our first attempt, when we were really novices at this PAF, we did a workshop, a ninety minute workshop, if you can imagine, on this issue, with a colleague, Craig, Dr. Craig Mason, who is very interested in what you are talking about, which we haven’t even dealt with yet, which are real causal pathway issues, where you have

direct and indirect affects. You have the direct affect of smoking and you just said smoking might also increase hypertension might increase low birth weight. So you have a direct affect of smoking, but you have an indirect affect through hypertension. Which is even different than what we are talking about here that is another fertile area for more research? Dr. Mason has done a lot of work on that, but there is more to be done parsing that out. That is a whole other question, but I would like to get back to interpretation because I think that…forget this chart for a while…just look at either one of these. To me, how much of a difference or not a difference do we want? I suppose here, this is controlling for poverty, if you were thinking about priority setting maybe, and I underline maybe, you want to say, maybe cocaine use, trying to eliminate drug abuse is going to have a bigger yield in terms of reducing low birth weight than smoking, but they are not that far apart, yet. I think that when you add this, this gets back to our original question about considering something modifiable or unmodifiable. Here you are saying, “Well, who should we direct money towards, what should we fund, smoking cessation or some kind of cocaine prevention and treatment programs.” Now you look at that and what is the interpretation of that? If you are really going to consider poverty modifiable, I think you have to say something about…even if you are being cautious and think that it is not poverty per say, but it is whatever those unmeasured, truly unmodifiable factors are that are associated with, that poverty is proxy for. We better find out what that is and we better start working on that, diverting resources to that if we really want to have a large impact on low birth weight. I don’t know if we want to go there, but to me that is what that pie chart says.

KRISTIN RANKIN: Any other questions?

UNKNOWNS SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: Right.

KRISTIN RANKIN: I will definitely add those slides back in now that I am not presenting my abstract, but the potential impact fraction would take care of that.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: We would have to go back and look at the actual data because my …If you think about using stratum specific estimates, both the relative risks are different and the prevalence is. Two strata is made up of a really small stratum and a really big stratum, in terms of the numbers. See that might make a difference too. Once you are not just using the adjusted measure you've now got new prevalence estimates, as well as new relative risk estimates, am I right or wrong.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: Right.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: If that is true, let's see if I'm right. I don't know if I'm right or not.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Sure.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: I think that is what…even though I think we should all unanimously tell Kristin not to do this, but that is what she was saying she was going to do tonight.

KRISTIN RANKIN: No I am going to do that tonight. You will get it tomorrow.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: Right. That is probably what it is going to be.

KRISTIN RANKIN: Keep in mind this one here also had the interaction term in the model. I can do these both without the interaction term. I wanted to do these examples using beta estimates instead of using the fully stratified approach, which was the first way that we taught this, fully stratified. When I move to modeling I generally created the model using backward elimination with the interaction term in it, not realizing that it was going to complicate explaining these things. So I will simplify that and we will go over it again tomorrow, so hopefully it will become more clear.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Uh huh. Right. It should be in addition to the slides not a revision.

DEBORAH ROSENBERG: Not a substitution.

KRISTIN RANKIN: Not a substitution, right. What else was I going to say about that?

DEBORAH ROSENBERG: Well was I right about in terms of the prevalence estimates in each once you are looking at stratified because of the product term, you had to use separate PJ's for that.

KRISTIN RANKIN: No the PJ's always stay the same.

DEBORAH ROSENBERG: They stay the same.

KRISTIN RANKIN: Because it is the proportion of cases in the…

DEBORAH ROSENBERG: Because they are already from the fully stratified approach.

KRISTIN RANKIN: Right. Go ahead.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: You could do that too.

KRISTIN RANKIN: Well that is a good point because when I said for my dissertation that I had stratified by an unmodifiable factor, like in this middle model, because we saw that interaction between smoking and poverty, at that point we could have stratified by poverty and done a pie chart for poverty equals yes and poverty equals no, and done smoking and cocaine and see how they differed. That would have been another way to handle that interaction term.

DEBORAH ROSENBERG: Right.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: Right.

KRISTIN RANKIN: Exactly.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: Right.

KRISTIN RANKIN: Because in that case, when you stratify you are definitely going to have different prevalence in each stratum, unless they happen to be the same, but you estimate them differently.

DEBORAH ROSENBERG: Again what we were saying makes total sense if we are going to call poverty unmodifiable. If we decide to call it modifiable, then we lose the ability to see a slice of the pie for poverty, if we are considering calling it modifiable, if we stratify in it, it won't be in any pie chart or table, whatever other kind of graph that we might want to make. So that is the choice.

KRISTIN RANKIN: Yeah.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Yeah, and in some respect it is kind of important to have your conceptual model to begin with and to kind of say "What is the story that we want to

tell," you can look at it several different ways and see what story makes the most sense. I don't know if I would advocate for that that is probably not a very purist view of things. If you do stratify by different ways, what story is actually capturing your population the best? I like this conversation because it brings up all the different ways that you could handle all of these different analytic challenges. Other questions or comments, we will follow up on this tomorrow. Do you know what time it is?

DEBORAH ROSENBERG: Yeah, it is four thirty, so I think we can introduce exercise two.

KRISTIN RANKIN: I will talk a little bit about what we are going to do for exercise two. I'm going to introduce it today and you can do as much as you want of it this afternoon, but we are also going to give some time for it tomorrow. I will explain what we are going…In this case, I've taken the Bruisey data, and this is beyond what Bruisey did with it. Bruisey did the adjusted PAF for each of the four factors and the summary PAF. I've taken the data and calculated all of these intermediate summary PAF's for all of the different possible combinations for these four variables. Now we are taking the example that I just did with three variables, extending it to four variables. I did all these calculations for you, but now what we want you to do id calculate the average PAF for each of the factors. So for age menarche, age at live birth, previous breast biopsies, and family history, you should be able to use all these values to calculate the average PAF for each one of those. So for four factors can anybody tell me how many sequences there are going to be? There will be twenty four different sequences. So

there are twenty four sequential PAF's for age at menarche, but remember that the first sequential PAF for age at menarche is going to come up how many times. How many times, you say four? You say six? Six is actually right. You have four factors, the three factors after; age at menarche could be arranged six different ways. Keep that in mind when you are doing your calculations because you have to factor that in. Do you have questions before we start? The other thing that I wanted to explain about the Bruisey data that we may be brushed over a little bit, three of those variables are ordinal variables. I still have the paper up there.