MCH EPI Conference

Using Population Attributable Fractions to Assess MCH Population Outcomes

December 8 – 11, 2008

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Exactly. Right.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Right and you know what? I'll definitely provide those resources for you. We probably should have included it. I was actually going to present an abstract all about the potential impact fraction at the conference later and I realized I was biting off more than I could chew with a six-week old. So, I pulled it and probably will do it next year.

DEBORAH ROSENBERG: While Kristin is away, Debbie, see actually I think this really is going to be interesting when we think about it's actually going to highlight if we wanted to report that 47% but to know that poverty is like half of that. Right? You know, even if you're not a smoker, you're not a cocaine user, the component PAF for just poverty was like 23%. Right? Yes, 23% or if we're using algebra ratios, 24%. So, what does that say? Just think about this because as Kristin said we're going to go on and go

through this example both ways. If you decide to consider poverty modifiable or if you decide it's not modifiable and what your perspective is and then how that would change what you can say about your results.

KRISTIN RANKIN: Okay, so are there any questions? We cleared up, I think that question. Yes, go ahead.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Preferable would be the one using the (inaudible) regression because you're actually directly estimating the relative risks, which is better than estimating the relative risks with the odds ratio. So, really the .45% is more realistic.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: I wanted to show that because it's just a limitation of there is a beta program that will let you estimate adjusted PAF and I'll show you that later but it only lets you estimate the adjusted PAF from the beta estimate from logistic progression. So, I just wanted to show you how to set that up. We wanted to reexamine the fully specified model because if you remember from the parameter estimates. (Inaudible) value for some of these interaction terms, they're not significant interactions. So, the benefit of modeling is you can build an appropriate model and this isn't necessarily appropriate if we know this interaction, first of all between the three different variables, smoking times

cocaine use times poverty has a P value of .11. So, we know it's not a significant interaction. We can really drop that from the model. We also know smoking times cocaine is .28 for P value. So, if we drop those two from the model we would get a reduced model with just the two interaction terms and then when you take out that three-level interaction term you see that the cocaine times poverty interaction term is also not significant. It's .09. So, we can also drop that one and this is like a typical like if you're doing backward elimination. If you're doing a full specified model, you're doing backward elimination in your model and there's lots of different ways to do model building. This might not be the way that you prefer but it's one of the approaches to eliminate the interaction terms in a backward fashion. So, if we've done that we just have one interaction term that's significant, .002 for smoke times poverty. So, now you can take these beta estimates from a more appropriate model and use them and we're back to the Excel file, letter C. Now, fill-in the betas here in the Excel file from the reduced model and come up with relative risks here for a better model and you'll see the summary PAF is .441. So, what modeling allows you to do and what using the beta estimates will allow you to do is to build a model that's appropriate for you beta and especially if you start having lots and lots of variables and use the beta estimates from that model to generate your PAF. I've show you both binomial and logistic again just to show you how the logistic progression overestimates it slightly. The other thing, if you think about this, if you want to control for five other variables that you're considering un-modifiable, which we'll get into a little more later but say you want to calculate this summary PAF but you want to adjust for other factors such as maybe region even. If you have different beta from different regions in your state, you want to control for that

variable or any other variables you can think of that would be appropriate. You could adjust the beta coefficients for those other variables. Use the beta coefficients from a model that controls for all those other factors and then you would have a summary PAF that would be for the summary PAF for smoking, cocaine and poverty, controlling for this whole other set of variables and that's why this is so powerful because you can use interaction terms, you can use compounders that are appropriate and use the beta coefficients from those. Are there any questions?

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: So, you should have the exercises in your packet, in your paper packet. You may have them electronically on your desktop now too. What we're going to do is take us through the article. I'm not sure how many people read it beforehand. We gave it to you as one of the readings but we know the reality. Maybe you didn't get a chance to read it as carefully as you would want to but this is the article that really presents the summary PAF and presents using the beta coefficients to generate summary PAF's and they give you enough data in that article because I was mentioning 72 unique cross classifications. They actually provide all that data in the article. So, we thought this could be a useful data set to use that's a little more complex where you can start generating some of these yourself. What we've done is provided on your desktop a file called Brewsy 1985 Data.xls, where I've actually copied all of the (inaudible) straight up from the paper into an Excel file so you can start working with it electronically. What we're going to have you do for Exercise 1 is calculate the adjusted PAF for one of the

variables. It's a breast cancer isn't it? The outcome for Brewsy? The outcome in that article is breast cancer. So, "age at first live birth is one of the variables. So, we would like you to calculate the adjusted PAF for "age at first live birth" and then the summary PAF for the entire risk system where there are four variables. One is dichotomous two or three-level ordinal variables and another one is a four-level ordinal variable. This is moving to a much more complex data set. Then, you can actually compare your answers. You have the answers in the article so that you can check your work, which I always find is helpful. They're using a modeling approach. So, you'll see in the paper you have beta estimates that you can use to generate summary PAF just like we did in these Excel files. So, I think it would be helpful probably to reread the Brewsy article before you start on the exercise. Then start to work through in Excel the calculations.

DEBORAH ROSENBERG: Just to point out in the Brewsy article this is actually case control data. So, this is appropriately odds ratio is not relative risk because the study design was match case controlled wasn't it? Yes, so frequently matched or something.

KRISTIN RANKIN: Yes, they actually used conditional logistic regression to develop the beta estimate, which we don't have to worry about because they've already given us the beta estimates.

DEBORAH ROSENBERG: So, I think just as a starting point because I was telling Kristin, I went and I tried to redo these exercises too to make sure I knew what was going on and just remember that the prevalence estimates are always based on the

empirical observed data. So, you've got in that Excel file the numbers of cases, truly cases in this case, numbers of breast cancer cases for every combination of the risk factors and you can actually use those to calculate what we've been calling the PJ's and what Brewsy calls the PJ's. Then you're going to have to do some data entry but it's minimal. So, by using the Brewsy article you're going to have to find the beta…you know the beta coefficients are there, enter them in the appropriate columns. So, you're modeling what Kristin just went through in Worksheet B really of the Excel file we've just been looking at. So that's the one actually that you may want to refer to as you're looking at how were the formulas implemented, etcetera. So that's it and I mean at this point, I hate to ask this question because I know what the answers going to be but I mean we could move on but I'm really thinking it's probably better for you to start this exercise before we get into any other complications and if we finish it before the 3:00 o'clock that we set in the afternoon that's fine but I just know from doing it myself, I think it's going to take you a little bit of time to do this and we really want you to have time to do this because until you struggle through the data manipulation yourself and make errors and go back and correct them and all of that it's just harder to assimilate what we're talking about. So, I think that we should let you have it and we'll break for lunch when it's times to break for lunch and take as much or as little time as you need and come back and we'll keep working on it when we come back from lunch and, of course, if you're really stuck ask us because it probably means everybody's stuck and we'll try and intervene.

KRISTIN RANKIN: If anybody has their own laptop and wanted to we have the files. If you wanted to work on your own or if you wanted to work in your pairs at your laptop that's fine too. However you want to do it. Just let us know if you need the files. If you have any questions, please ask us right away. We want to help you through this if you need.

KRISTIN RANKIN: Okay, I think wherever you're at now come to a stopping point and we'll start talking about it. I think as some of you discovered there are several different ways you could kind of go about this. I'll show you one way, which is actually a way that Deb did and it's not how I would have done it but its definitely right. Well, it's actually an easier way. I always go the hard way around and I was probably helping you all to go the hard way around but it's good if you went the hard way around it means you really understand it. So that's good. We could start with the summary PAF because I think that's more intuitive. The first part you needed was that PJ, right? So, it's a proportion of cases that are exposed in each of these strata and to calculate that…

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Yes, um-hmm, just a second. Does that help? So, in order to get the PJ you want to take the number of cases. So, you can see this formula here is E2 over 948, which is the number of cases, right? We just put the 948 in the formula and just dragged it down but the other way you could have done it was to sum across Column E,

which is all the cases. We have the sum down here. It's in E74. So, this is this little trick that I was talking about that you can do. You could do E2 over 74, dollar sign E, dollar sign 74. You have to put a dollar sign in front of the column and the row. Then you can drag that down and it keeps the denominator the same. So, you have E74 there but it changes the numerator to that row.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: In front of the row, right. So, I just had to put E$ and 7 before, right?

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Yep. I over did it. Again, the hard way around I went. I got the right answer. So that's how that works. So, is everybody okay with getting the PJ? So, then the next thing we wanted to get for the summary PAF was the strata specific relative risk and we're struggling with these terms. Deb just suggested maybe we call this the component OR. I keep saying relative risk but in this case it is odds ratio. I'm sorry about that. I'll call it odds ratio from now on. We thought maybe call it the component OR because there's two different types of strata specific OR's that we're dealing with here. So, this is the tricky part but basically what this odds ratio is comparing to a common reference group. So, let me hide so we can get back to the raw data, we'll hide all of these rows and these columns. So, in this case, actually I didn't want to hide those because it has all of the information in it. Okay, basically how Deb did this is she took

the odds ratio that was in, I think it was in Table 5. She used the odds ratio instead of the beta estimates from Table 4. I was thinking of it in terms of beta estimates but you could also have used the odds ratios that are in Table 5 because those are just really the beta estimates multiplied by the level of the variable and exponentiated. So, you could go the odds ratios and then go back. So, in this case, the adjusted odds ratio for each of the variables is here. So, let me hide these. Basically, we'll give you something where we've figured out the adjusted for all four variables. We'll give you that and the answer key. We'll just go over one of them today. So, if you put all of the odds ratios in this Excel file it would be one across the board for all the first ones because this is that common reference group is zero, zero, zero, right? Then, as you can see we'll use age at last live birth. This is the 1.3 and then 1.7 depending on the odds ratios that were given in the table. So, then if you want to apply those in the formula, you would just sum across all of…wait how did you do this? Now, she's taken the natural log of the odds ratio. So, going backwards and getting the beta coefficient for each factor for each of these and then she sums across them and exponentiates the sum of them. So, you start with the odds ratio, take the natural log, right?

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Well, it's not cheating. You did the same thing as if you would have gotten the…but you have the beta estimates here. She added across these and exponentiated them to get the common OR . The key here is that this odds ratio in the first row is one and all of these are assuming that you're dividing by this one. So, when

you exponentiate the sum of all of these you could have divided by Y dollar sign 2 and that's really what that's doing. So, your common reference group is here. This odds ratio is compared to this and this odds ratio is compared to this, etcetera, etcetera. So, when you have all of those odds ratios you can do…and we have both methods here. So, for Rothman you could do the PG, which is in column G times relative risk minus one over relative risk and that relative risk in this case is the component. The odds ratio in this case is the component odds ratio in this column here. So, then you just drag this down and you have all of the…these are all component PAF's and you sum them to get the .5465. You could have applied the Brewsy formula and taken the PG, which is in Column G, divided by the relative risk.

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Right. So, with Brewsy these aren't component PAF's now. All of these individual values don't have any meaning until you apply this and you do one minus the sum of all of these but they're equivalent. You could have done it with the formula. Was that enough of an explanation or were there questions? The other way to do this, the way that I think I thought about it and the way that was kind of equivalent to how I did it in Worksheet B and Worksheet C and the supplementary files was to take the betas from Table 4 and apply the beta to each of the codes and then exponentiate to get the relative risk. The same difference though. Was everybody able to get the summary? Okay, great. So, for the adjusted, the easy way out with this was to take the relative risk from Table 5, as Deb did here. So, let me show you. This is the adjusted for

age at first live birth. So, if I move this over here I want to show you. I don't know how to do that. Well, at least I can show you the value for age at first live birth here. Okay, this kind of gets at it. So, for the age at first live birth adjusted you have the odds ratios from Table 5 and so she knew because the value is zero here, the odds ratio is one because that's the reference group. Now, when the value of age live is one, it's 1.306, which is from Table 5 and then you go all the way down and you get 1.7, 2.2 and then you see it repeats again. The way the data are sorted and unless you've resorted your data it repeats again so that there's a reference group again for age live of zeros though you go back to having an odds ratio of one and you go 1.3. So, this was the easy way to do it. I kind of guided some of you through the harder way but if you understood it that's good. So, then if this is your odds ratio, you're applying it to the PJ and apply the Rothman or the Brewsy formula in the same way you did for the summary but now since you're working with a different relative risk because it's the strata specific relative risk for comparing. In this example you're comparing Row 8 to Row 2 because this is age live one with PBB at zero and family at zero, where age live is zero and PBB is zero and family is zero and there's one other variable in the beginning. So, you're comparing wherever age live is one to zero and then you're comparing two to zero and three to zero within the strata of all the other variables. Does that make sense? So, then when you sum across it you get the .3443, which is what's published in the paper for the average PAF. Now that was the easy way to do it. The other way you could have… What?

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: I'm sorry the adjusted PAF, right. We didn't get to average yet. The other way you could have done this is because you've calculated all of these odd ratios already, these component odds ratios, you could have taken each stratum and divided the odds ratio for that stratum by…

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Yes, so basically what I've done here is if you're in Row 8, if you didn't resort your data in Row 8, you would be comparing the component OR for Row 8 to the component OR for Row 2 because basically in Row 8 the value of age at last live birth is one and in Row 2 the value is zero and then all of the other values are the same for those two rows. So, you just kind of make sure you know what the reference group is and then I was able actually to drag that all the way down, drag that formula down because you're going to have the same relative risk, odds ratio. Then, I could drag it down when age at last live birth was 2. I dragged that all the way down. Age at last live birth was 3 and drag that down and then I did figure out that I could copy this entire thing and when it starts over again, where age at last live birth is zero again, I could just paste this until it became 3. I dragged it and pasted it. Am I losing you? Anyway, it wasn't as hard. You didn't have to type in the formula each time. It wasn't a very good explanation. It's really hard to explain an Excel file I'm realizing. So, was everybody able to get the adjusted or if not, where did you get stuck?

UNKNOWN SPEAKER: (Inaudible)…

KRISTIN RANKIN: Yes and the reason that you could ignore the other three variables are because they dropped out of the equation. Are you going to go write this up? Because you're comparing age at last live birth one versus zero within the same group of the other variables, they all dropout because they're in the numerator and the denominator when you're doing a relative…

DEBORAH ROSENBERG: In fact that's the definition of an adjusted, right? If you think about when we model and you go E to the beta zero plus beta one and I'm going to make it simple. You know, dichotomous variables. Let's say this is one and we're comparing one versus zero adjusted for everything else. Let's say you have beta two times one. There, I'll just do it as two variables. This is the comparison, one versus zero and saying you're adjusting says that holding everything else constant. So, here I am holding this constant and I just end up with then E to the beta one times one. So, E to the beta one here, this drops out because I can do this when beta two is zero, it's zero, zero. It doesn't matter and that's the whole point controlling for whatever two was this is the adjusted as opposed to what we were doing before, which is whether we want to call them component OR's or something where I'm changing the values in different strata for lack of a better…let's say…plus beta two times one and then maybe in the denominator plus beta one times zero plus beta two times zero or something like that. That's not an adjusted odds ratio. We haven't held anything constant. It's a whatever,

strata specific or a component odds ratio or relative risk. So, this is the whole meaning of adjustment is holding everything else constant.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: Right.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: Right.

UNKNOWN SPEAKER: (Inaudible)…

DEBORAH ROSENBERG: Right and what I'm reinforcing is it's funny that we all do this kind of regression, forget PAF and we say these things without really thinking, you know, the effect of smoking regardless of like you're saying. We say those words, well that's what it really means and so this just reinforces even when we go back to doing regular old modeling for ratio measures what it is that's happening there.

KRISTIN RANKIN: Somebody pointed out that that's why you could have used these odds ratios in Table 5 for your adjusted odds ratios to get your adjusted PAF but you needed to go back to the regression coefficients in Table 4 to get the summary because you didn't have all of those different combinations for all of the different strata of the 72

different rows. So, do you feel like you understand that enough to go on that you could calculate a summary PAF because these are the building blocks for what we're going to be doing?

DEBORAH ROSENBERG: I think we're going to go on and we'll break when we're supposed to break, which is 3:15 and one thing, just I was joking with someone here but it's actually not a joke. I keep saying to Kristin this is why we need Prock Rankin and SASS to do this. I mean who wants to do this in Excel? We don't want. We want to tell Sass what to do and let it do it or Prock PAF. We need Prock PAF. You'll see as Kristin said in strata, Kristin will be showing you AF logit but it's again, it hasn't gone beyond doing the adjusted and as she has said many times using logistic regression. So, it's still not a fully developed way even in strata to do what we want to do. So, we are going to go on. We've done the summary PAF. If we want to call them component PAF's based on those component relative risks or odds ratios and adjusted, which we said is the first sequential PAF. Now, we're going to talk more generally about sequential PAF's and we're going to start once again from a really nice, simple trivial little example back to our smoking and cocaine and then we'll again move into a more complex version and using modeling that Kristin will talk about after the break. So, let's talk about in this little small risk system how we'll calculate sequential PAF's. So for this risk system with two variables there are two sequences. We can eliminate smoking first controlling for cocaine use and then eliminate cocaine use or we can eliminate cocaine use first and then eliminate smoking. Within each of those sequences there are two sequential PAF's. So, let's first do the sequential PAF's for eliminating smoking, controlling for

cocaine use. This is the first sequential PAF. So, this is what we already calculated. This is the adjusted PAF, smoking controlling for cocaine and that's this notation just "smoke given cocaine", the S bar C is just "smoke given cocaine" and if you remember from before we had calculated that as the .076 or 7.6%. This is what we haven't done. What's the second of these sequential PAF's for eliminating cocaine use after we've already eliminated smoking and how do we do that and actually it's sort of a…I don't know, sometimes Kristin and I think of this as cheating in a way but I think it makes theoretical sense as well. So, we've calculated a summary PAF for the smoking cocaine risk system. We've calculated a first sequential or adjusted PAF eliminating smoking first, controlling for cocaine and if we take the difference, the remainder of those two is what's leftover for when we eliminate cocaine. So, in this case we have the summary PAF was calculated as .16 or 16%. We calculated the first sequential PAF for smoking adjusted for cocaine as the .076. We take that difference and we get the PAF for eliminating cocaine second after smoking is already eliminated. So, we've just solved for the unknown. We have two of the pieces of this calculation. Obviously you can see that and you will see that when we get to more than two variables it gets a lot more complicated than this. So, now we can do it. How about the other sequence? We can eliminate cocaine first. So, now we have C given S, cocaine given smoking, this we also calculated before because it's the adjusted or the first sequential PAF and we had gotten 0.099 and now we do the same thing. We take the remainder. We know what the summary PAF is. Again, 16%, the same risk system and now we're going to calculate the difference between that summary PAF and the first sequential or the adjusted PAF for cocaine. So, we have the 0.16 minus 0.099 and we get .061. So, now we have two

sequential PAF's for each of the two possible sequences given this very small risk system that just has smoking and cocaine. So here are the pie charts and we said that by definition we just solved for the unknown. It's all about the remainder from the summary PAF. So, by definition, the sequential PAF's within the two possible sequences do sum to the summary PAF. So, we have the smoking first and cocaine use first are the two sequences and then you can see the different partitioning of that .16, the summary PAF, the left-hand pie. When smoking is first, we've got the .076 for smoking, .084 for cocaine and when we do cocaine first we have .099 for cocaine and .061 for smoking. Well, which sequence is right? And again, you can imagine how fast the number of sequences escalates as we add variables to our model, to our risk system. So, this is very simple. We only have these two pies to compare. They sum to the PAF but, which again, which one do we choose. We still don't have one single measure of a PAF for smoking and one single measure for a PAF for cocaine. We can take our choice in this case between two sequences as we move on between three, four, ten, twenty, thirty, a hundred different sequences depending on how complex our risk system is. Kristin, when you did the work with NSCH do you remember for your final model how many possible sequences you had? I mean was it in the order of dozens or…?

KRISTIN RANKIN: (Inaudible)

DEBORAH ROSENBERG: It was more than that. So like a hundred or something?

KRISTIN RANKIN: (Inaudible)

DEBORAH ROSENBERG: More than that. You didn't have that many…what did you have maybe six or seven variables in your final model?

KRISTIN RANKIN: (Inaudible)

DEBORAH ROSENBERG: So, I mean it's not as though you had twenty-five variables in the model. You had relatively 720 sequences. So, it really gets complicated fast and so again that means you have for each sequence a different PAF that you could theoretically report for any given factor and that's not very satisfying. We want what happens on average, which is why we're going to calculate an average PAF. Again, we're going to use the words here, regardless of. So, regardless of the order in which smoking is removed from a risk system or cocaine is removed from a risk system, on average, what's the PAF? What's the proportion of low birth weight that can be attributed to smoking or cocaine? So to calculate the average we kind of have to rearrange the sequential PAF's that we were just talking about because we were going first sequential for smoke and then remove cocaine or first sequential for cocaine and then remove smoke but we want to combine all the sequential PAF's related to a given variable together in order to create the average. So, now we're going to go eliminating smoking first averaged with eliminating smoking second. So, we're crossing sequences here. Or, eliminating cocaine first averaged with eliminating cocaine second and, of course, again so you can see average with eliminating it third, eliminating it fourth,

eliminating it fifth, eliminating it sixth and there are multiple sequential PAF's similar ones, non-unique ones associated with those cases and you'll see examples of that as we move forward. So, for this smoking cocaine risk system what you see here are the calculation of the average PAF's. So, we've got the first sequential PAF, which is the S given C, smoking controlling for cocaine and then plus the PAF's for smoking after cocaine has already been eliminated from the system and we have two. So, it's a simple average of dividing by two in each case. The second one is for cocaine. So, we have the .076 for smoking plus by .061 divided by two and we get 7% for smoking. We have the .099 averaged with the .084 dividing it by two and we get .09. Now, we have one single estimate for each of these two factors and here's the pie chart. Now, we still add up to sixteen the summary but it's just the .07 and .09. As we move through this then it's that summary PAF represents the total for the entire risk system however we define that risk system and then it's a matter of how many different ways can you skin a cat? How many ways can we partition that summary and which way is most satisfactory and we're beginning to think that getting to this average is the most satisfactory even though you can do all those components like we've done or you can't use the adjusted aren't a partition but you can look at sequential PAF's, whatever but the one that probably has most utility is this average because it's single for each factor and finally we don't have that area of the pie where we still have overlapping instances of women who are using both. We've managed to tease them apart and repartition the summary by using this technique of averaging. So, I don't know. Do you want to start this Kristin before the break? It's ten to three. So, actually, we could go on for another twenty minutes or so. So, did you want to start this?

KRISTIN RANKIN: So, I mentioned before that with the three factors smoking, cocaine and poverty, when we start thinking about them in terms of sequential and average PAF's, we want to think about the modifiability and un-modifiability of these factors. So, I did this as kind of a scenario and I want you to make the decision about how we go forward but we're going to do it both ways anyway. I just wanted you to know what ahead of time. The scenarios that you're asked to prioritize spending for interventions that target the high rate of low birth weight in your jurisdiction. You have a data set and its the same one we've been working with, with relatively reliable data on smoking during pregnancy, cocaine use during pregnancy and poverty levels, which would be a nice data set to have, right? We don't have a lot of income data on what we have but this is kind of a…that's why we're using fake data. We would like to use one of the methods you learned about to calculate a PAF for the factors. We're going to use a modeling approach because as we learned within one model we can differentiate between those factors considered to be modifiable and those considered to be un-modifiable and this doesn't change the model but it does change the calculations of the PAF that we decide to calculate. So, first of all, we have some decisions to make. Would you consider each of these un-modifiable or modifiable? So, smoking is there any question or would we call it modifiable all the time probably, right? Or, I mean, I know there's been a lot of smoking cessation programs and people feel like they're banging their head against the wall because it doesn't…you know, they're not always effective but I think for the most part we're hoping that we can modify that behavior

eventually for everybody, right? So, cocaine use, do we consider that modifiable or un-modifiable?

UNKNOWN SPEAKER: (Inaudible)

KRISTIN RANKIN: Okay, what about poverty?

UNKNOWN SPEAKER: (Inaudible)

KRISTIN RANKIN: So that's a really good point, right? So, it kind of depends on where you're sitting. If you're sitting within the state health department you're thinking we don't have programs to make a difference in terms of poverty but if you're sitting maybe at a higher level of government or in the federal government maybe thinking about or in other areas of government and thinking about maybe there are ways to do this. So, it really depends on your perspective in terms of where you're sitting and also…

DEBORAH ROSENBERG: Can I make a comment?

KRISTIN RANKIN: Sure.

DEBORAH ROSENBERG: Because this follows up on what Wanda was saying, which is that part of me, I'm of two minds because part of me, exactly what was just said. You know, it's not fair to hold people accountable for what they don't have control over. So,

from like the healthcare professionals point of view it's hard to ask the health department to eliminate poverty. On the other hand I feel like if we don't calculate a PAF for poverty, it's really a copout because it's really saying this is where we really need to put resources. It's great to have the smoking cessation programs and great to have the substance abuse programs, great to have all of those other things but ultimately until we really focus on the bigger social questions, we're not going to eliminate low birth weight or infant mortality or any of our common, you know, and so part of me wants to force us to make something like poverty a modifiable factor because it challenges our allocation of resources, etcetera and our focus.

UNKNOWN SPEAKER: (Inaudible)

DEBORAH ROSENBERG: Right.

KRISTIN RANKIN: That's a very good point. The other issue with something like poverty that a lot of purists, epidemiologists would say, "Is there a causal link between poverty and low birth weight?" So that's where we kind of get into these causality issues. So to say that one proportion of low birth weight is attributable to poverty.

UNKNOWN SPEAKER: (Inaudible)

KRISTIN RANKIN: That's one very good point that your reference group and in this case we're calling it below the federal poverty level, above the federal poverty level. So,

above the federal poverty level is your reference then maybe that's not the group that would get you to be disease free but we're going to go over that in a lot more detail tomorrow when we're talking about choosing your references when you're using this method. In a case where you have smoking and non-smoking, unless you're looking at dose of smoking you have a pretty clear yes/no but in a case like poverty where you're setting a line like that your choice of reference is really important because that may impact the value of your PAF or something like in my dissertation I used hours of TV watching for kids. Where are you going to set the reference level at? Where do you want your standard to be and then how do you want to calculate your PAF from there? So, that's a really important point that we're going to get to tomorrow but the causality issue too is that poverty probably comes with a lot of other factors that are causally related like low birth weight. So, what are we estimating here?

DEBORAH ROSENBERG: That's what we were saying about race/ethnicity. It's the same sort of thing. It's a marker for a lot of things that we're not measuring and so then the question does become do you want to highlight that or not? I don't know.

KRISTIN RANKIN: Exactly. So, we're bringing up some more questions than answers in this case but mechanically I'm going to show you how to go through it both ways so that you know how to calculate PAF's for two factors controlling for a third factor or the three factors across all three factors but we wanted to bring up that conversation because there are a lot of different ways you can handle that. Then what type of PAF do we think is most appropriate for this analysis? I think we've already kind of given away our bias

that we think the average is probably the most useful because you can actually compare several factors. So, I wanted to just show you. I wanted to start with the data and show what we have here because we haven't seen this poverty variable yet. So, we have smoking and cocaine, the same distribution as before, the same data set. We have our 700 low birth weight cases but then we show here that it's actually a pretty high risk sub-population that we're working with here. So, in this fake data set we have almost about 42% of the women in our data set are in poverty. So, just so you know, you have that perspective. So, when we're looking at the results you understand why. So, the prevalence of poverty is high overall and is high among the cases as well. It's over 50%. First, we're going to go through considering poverty as un-modifiable. So, right now, we're handling poverty as a confounder in the relationship. So, we're adjusting the other estimates for poverty without calculating PAF for poverty. There is still the two possible sequences that Deb mentioned when she was talking about just smoking and cocaine. We have eliminating smoke first and then cocaine and then eliminating cocaine first and then smoking and within each sequence there are two sequential PAF's. So, this doesn't change it just that these estimates are going to be adjusted now for the effect of poverty. I gave you some Sass code again here. This is for looking at that total cross classification of these three factors for the cases and then generating beta estimates. So, we're going to use the modeling approach here and this is from binomial regression to get relative risk directly. I used the same model that we came up before, this reduced model with just the smoke times poverty interaction terms since that was the only one that was significant. So, we're back to the supplementary Excel file. I'm want to show you how we went ahead and calculated sequential and

average PAF's. Oh, I'm sorry. We're just doing the first sequential and the summary PAF's right now for when poverty is un-modifiable. This should look familiar but the difference is here that the beta estimates have changed because…I guess the beta estimates are the same as that example that we did before. So, the summary should be the same as that last example. Oh, I'm sorry that will be the other file. So, anyway, we're doing the first sequential PAF's here for smoking and cocaine and remember these are controlling for poverty. So, in order to do that it's really similar to what we just did in the exercise. So, in this case, I was applying the beta estimates. So, in order to get the relative risk for smoking, see in Column I, I exponentiated the code times the beta coefficient for each factor. So, it was smoking one times the beta for smoking, coke one times the beta for coke, poverty one times the beta for poverty and then smoke times poverty times the beta for that interaction term and that's how I got the relative risk for that strata. That's not right is it? I didn't include poverty because that's what we're controlling for. So, what I'm doing for the relative risk for just smoking, I'm doing A8 times E8. So, the beta for smoke times the code for smoke plus the beta for smoke times the beta for poverty times the interaction term. So, what I want to do for this first sequential PAF for smoke for that relative risk, I'm doing smoking within every category within every strata of the other categories. So, you can see these are the references here. I did this a little bit differently than we did the exercise. I'm sorry, Deb, can you help out here? I'm not remembering how I did this. These are the adjusted relative risk for smoking actually but I'm trying to remember how I did this. So, I exponentiated the A8 times the E8, the beta for smoke plus the beta for smoke times poverty times the interaction term for poverty.

DEBORAH ROSENBERG: (Inaudible)

KRISTIN RANKIN: Right, it's complicated by the interaction term but these end up being adjusted relative for smoking. Is it clear why that these are the adjusted and then it actually becomes one because these are the reference groups.

DEBORAH ROSENBERG: (Inaudible)

KRISTIN RANKIN: Turn your microphone on.

DEBORAH ROSENBERG: Going back to the way I've been drawing it, I mean theoretically the minute you have an interaction term in the model, technically you don't have an adjusted anything. It's strata specific. That's the thing. So, technically it's not an adjusted. It's using the strata specific estimates.

KRISTIN RANKIN: That's why I have two different relative risks here. I have a relative risk within poverty equals one and a relative risk for poverty equals zero because there's that interaction.

DEBORAH ROSENBERG: That's the other thing we all just sort of forget because we're so used to…I mean the minute you have a product term in the model, the variables involved in that product term, there's no such thing as, you can't get an

adjusted relative risk or odds ratio by definition by putting that product term in you must be using strata. It's saying there is not homogeneity across the strata and by definition you've got different odds ratios.

KRISTIN RANKIN: So, let me go over this one more time so that it makes sense, if I'm remembering how I did this. In cell I8, what we're doing is exponentiating smoke times the beta for smoke, then smoke times the beta for poverty. I didn't include cocaine or poverty because they drop out because we're calculating those within the strata of the same strata as coke and poverty because we're getting the stratum specific adjusted relative risk for smoking. So, this is for smoking within the strata of poverty equals yes, which is 1.101 and relative risk for smoking within the poverty equal zero is 1.776.

DEBORAH ROSENBERG: You know, Kristin, maybe the better way of thinking about this is there are no adjusted relative risks here but what you're going to get is an adjusted PAF but there's no adjusted relative risk here. They're all stratum specific because of the interaction term.

KRISTIN RANKIN: Right, except for that when you go to cocaine.

DEBORAH ROSENBERG: Yes because cocaine is not involved in the interaction.

KRISTIN RANKIN: Because cocaine is not involved in the interaction. We just have one adjusted relative risk for cocaine, which is 4.209. So that is the exponentiation of

cocaine times the beta for cocaine and you see where the reference group is here. It's where cocaine equals zero. So that's where the relative risk for cocaine is one and then you have the PJ's as usual. So that's D8 over D16, which is the 700 and the PJ's are the same as they have been throughout all of these examples. So, then when you apply the Brewsy method and sum you get a first sequential PAF of .0735 for smoking. So that's slightly different. I can't remember what was it before you control for poverty? It was .076. So, it is reduced a little bit you see because after you've controlled for poverty you're going to have a reduction.