Missing Data in Longitudinal Studies: Mixed-effects Pattern-Mixture and Selection Models—2010 MCH EPI pre-conference training workshop

DON HEDEKER: So invariably in human studies in which we have repeated measures across time we're going to have missing data. It's a fact of life. It's not going to be the case that we don't have missing data, so we want to think about it carefully and do some appropriate modeling. And I'll be talking about in particular stuff that comes from a paper that I worked on in *Psyche Methods* awhile ago, "Application of Random Effects Patterned and Mixture Models for Missing Data in Longitudinal Studies" and this has also been done in my book in chapter 14 to a great extent.

Now we're going to look at this same dataset we've just looked at, but I'm going to treat it now as continuous just to make that part of the modeling a little bit easier, but clearly, everything I described from here on in could be also done with a dichotomist model as well. So we're going to look at this schizophrenic study in which we've got the severity of illness measured across time and again, as I mentioned earlier this aspect, the completion rate was quite different for drug and placebo was at first to me a little bit worrisome because I thought placebo 65%, that's not that great. 80% for drug, that's not so bad, but 65% for placebo that's not great. Is that playing some role in the analysis of this data? That is what I want to explore.

Now I should say it seems to me in many studies like this where you have intervention and control this idea that there is more missing data in the control situation it's pretty common really. They just get typically less attention and so therefore they've got more probability of dropping out it would seem, so this 65, 85 is perhaps not that unusual really.

Now let's take a look at some descriptive statistics. This variable now considered as a continuous response across time, so here I've got the placebo and drug group at the four dominate time points, zero, one, three, six and you'll notice that both of them are starting off means slightly above five , so what does that mean? In terms of the outcome everybody is starting of on- not everybody, but the average is between markedly and severely ill.

And then across time you can see the placebo group improves on average to be four and a quarter. The drug group goes down on average to be about three, so in terms of the averages then the placebo group is improving on average moderately ill. The drug group is going to mildly ill on average. So again, it seems like there might be some drug by time interaction. They start off the same, so there is no drug difference at time zero and then they diverge. The placebo group goes down to some extent. The drug group seems to go down to a greater extent.

Now on the continuous metric we can calculate what is the pooled standard deviation at each time point and you see exactly what we saw with the **** dataset. It grows across time. People start off in this study. They're all to some extent across time some improve, some don't improve. That variation just increases, likewise the correlations. Again, I've done the pair wise and the list wise deleted versions. You see exactly what you saw with ****. As you go away from the diagonal the correlations diminish. Within a lag they tend to increase. There is not a great deal of difference depending on whether I calculated them with the list wise or the pair wise. It doesn't seem like it at least, but clearly, assuming a random intercept model, which says compound symmetry is not- doesn't seem to be logical. The variances are not the same. The correlations not the same. They systematically differ from that in very obvious ways. The variance grows across time. The correlation diminishes as you increase the lag.

Now here is the plot of the mean structure across time and this is what I had to work with, with a psychiatrist. We looked at these means across time and again I saw this and I said well I can handle that with having a time and time squared in the model. He wasn't so fond of that idea. Now when I was having continuous Y the first thing I tried was I tried transforming the dependent variable because in statistics that is what you often do is you transform the dependent variable, so I tried that. That didn't seem to give me any better results in terms of the Y, X relationship, but again, when I went to the square root of week I got a rather good linear fit for the means across time. So that is what I'm using here for my affect of time. It's not a time and time squared or piece wise linear kind of model. It's a linear model, but in terms of this played with X axis, the square root of week.

We fit a random intercept and trend model just a little while ago with a binary dichotomized version. Here let's fit this with the random effects version. Now again if you look at this equation I want to point out that all that marginalization we had to deal with that admittedly was rather complex. We do not have to deal with that here when we have continuous Y because on the left-hand side we have the dependent variable. We don't have this logit of the probability. We don't have this nonlinear function, so all of that marginalization stuff that admittedly is kind of tough it does not play here when you have continuous Y. You don't have to worry about that. So plugging in zero here, zero here gives you the mean response. When you have a nonlinear link it doesn't give you- it gives you the response for the means Y that is not equal to the mean response. Here it is. The two are the same, so you don't have to worry about that. So analyze all your outcomes as continuous and you won't have to worry about this, right. If life could be so simple.

So this is a random intercept and time model just like what we saw with the binary, but here I'm treating it continuous. Again I'm using this metric of square root of week, so let's run this and for the heck of it I ran it on those subjects that did make it to the end of the trial. There was 335 subjects that made it to week six and incidentally that is what I'm defining as completer, not that they had complete data across time, but that they were measured at week six, so completer could be a person that has maybe not measured at all time points, but they're measured at week six. All subjects, all data from all subjects. Now let's look at these results. **** intercept, in this model that is the mean of the placebo group at baseline, 5.2, 5.3, so it's pretty similar in terms of these two models. Drug is the drug difference at times zero. At times zero is there any

difference between the drug and placebo patients? Well the all subjects' analysis says there is virtually no different. The completer analysis it's tending to differ, be different, so it's not .05 admittedly, but it would say that the drug patients that completed are a little bit more severely depressed at baseline than the placebo patients that completed. It's tending in that direction. It's not .05, but that's the direction of it.

Time effect, this is the response for the placebo group. They both agree in terms of their conclusion. They're both significant, but note that this estimate is a bit more negative than this, so among the completers the placebo completers seem to have a more negative slope than among all of the placebo patients because this combines completers as well as incompleters. Again I don't want to make a lot of it. It's a similar conclusion, but that is sort of the direction of things. Drug by time kind of works in the opposite manner, among the completers the drug by time difference is a bit less, but among all subjects. Again highly significant in both cases, so not usually different, but a little bit to raise your eyebrow and make you wonder.

Now of these two which analysis would you rather go with, the completer analysis or the all subjects? I mean in an intent to treat analysis you would want to analyze all subjects that are randomized, so the completer analyses could suffer certainly from a selection effect. Maybe the people who don't complete the study are systematically different than those who do and different and this study was designed to measure all subjects that were randomized. So the all subjects' analysis would really be preferred here. I put this on in a sense to just as comparison just because I was kind of curious about this because sometimes people do analyze or in the old days they certainly did this kind of thing. They only looked at people who made it to the end.

Now the additional things are the random intercept and time effects, the co-variation. There is not much difference there and I don't want to go into that in super great detail. We've kind of talked about that maybe enough, so that is what we're seeing there. We're seeing significant **** in the starting point as well as the trends and not a whole lot of association between the intercepts and time, in fact, virtually none. Yes.
Question: Can I just verify, so the time and parameters the time here means it's among the placebo group?

DON HEDEKER: Correct.

Question: So the **** one when ****.

DON HEDEKER: Per square at a week for placebo only.

Question: Only for the placebo.

DON HEDEKER: And the reason for only placebo only is because I have the interaction in the model. If I don't have drug by time interaction in the model-

Question: It's ****.

DON HEDEKER: It's really for both groups, but because I have this in the model- Remember whenever you have interaction in a model that modifies the meaning of the main effects. When you have drug by time in a model drug represents the drug difference when time equals zero. Time represents the time effect when drug equals zero and drug by time says how does that change, so it's really important and admittedly a lot of emails I get is precisely about this question. How do you interpret the main effects when you have interaction in the model? So it's an important thing. All you got to think about is like you got drug by time in the analysis, so drug represents the drug effect when this equal zero. Time equals this effect when drug equals zero.

Question: Could you interpret the interaction ****?

DON HEDEKER: Right and the drug by time interaction then would, so let's say this is the effect of time for the placebo group. This is at how much different is it for the drug group.

Question:****.

DON HEDEKER: Right, you add them together to get the time trend for the drug group. Exactly right. Or another way to think about it, this is the drug difference when time equals zero. This is how much the drug difference changes per unit of time, so it gets more negative per unit of time. It starts off being zero. The drug difference is virtually zero times zero. The drug difference when time equals one would be this plus this. When time equals three it would be this plus the square root of three times this, etcetera, etcetera. Does that make sense? So the interaction is telling you how does the main effect change that's why if the interaction is zero then typically you remove it because it's basically saying that these main affects don't vary and therefore you don't need the interaction.

So here is the fit of the means across time for these two groups with these four estimates from the mixed model and all I did to get these lines were then to plot in the different values for drugs, zero, one, different values of time for the drug by time interaction and do that little bit of arithmetic to gain this line and this line. You see the placebo group is going down a bit. Drug group is going down a bit more, so there is in effect of placebo, but you can see it's not tremendously large. It's significant, but clinically is it important to go down from 5.5 down to about 5? Not so much maybe. Drug group is going down on average from 5.5. Down to the end of the study they're down to like 3, so that would seem to be rather more improvement, clinically important.

So placebo group does go down. I don't know whether you would argue that is clinically important or not. It's statistically significant, but going down from five and a half to about five on average that doesn't seem to be that big of a change. I think I got that right or is it going down to four and a half? I don't know. Did I mess that up? Let's see. Here is two. That is three. Yeah, maybe it's four and a half. It might be four and a half. Maybe I got that wrong. I need to go to the optometrist obviously.

Now let's talk about the missing data issue and incomplete data models. Now a basic idea in the literature on missing data is the separation of the dependent variable vector in terms of the

portion that is observed and the portion that is missing and to realize that whether a person if observed or not at a particular time point that is like a random variable in and of itself and so that is a variable we can think about in our potential modeling of the longitudinal data. So that is the variable R. It's an indicator of whether or not a person- an observation is observed or not.

Now Ruben [ph] in I think '76 published these missing data mechanisms that he called. You may not agree with his terminology, but they've been around for like 25 years, so you're not going to change them. Some people have tried, but largely been unsuccessful. What Ruben called missing completely at random is truly what I think the layperson would understand as missing at random. It's completely at random. What it means is that the probability of being missing can depend on X, that is covariates, but cannot depend on Y and this is what GE assumes in its estimation. It assumes this version of missing completely at random where the reason the people are missing are not- has nothing to do with the dependent variable, either the part that was observed or the part that you have missed. It can depend on X though. To be honest, this is a special case of missing completely random where the probability of being missing can depend on X covariates, so conditional on covariates R is independent of either the observed or missed portion of Y. Now what would be an example of covariate dependent missingness? Well let's go back to this slide on this study. 65% and 81% is group something that is going to be in my model as a covariate, so in and of itself the fact that placebo has more missing than drug that's not in and of itself a problem. In other words if the missingness is only because of this that would be consistent with missing completely at random, covariate on missingness. At this point we don't necessarily know that, but just this difference, when I was initially worried about it, in and of itself that is not a problem necessarily because we have group in our analysis. Group is in our analysis, drug group, so the fact that one group has more or less missing data as long as group is in the analysis that is not necessarily a problem if that is the only thing that explains missingness. Now that is kind of a big if, but realize that that's not in and of itself a problem. Yes.

Question: I apologize if I'm totally ****. I thought that we were looking at Y, which is the measurement of illness, so I understand that you're **** covariate and but that is where the people that **** group and the people are **** Y, so how is it that is not ****? **** next slide. **** random. **** conditional that it's on X and not Y, but we're actually looking at Y.

DON HEDEKER: Correct, we're looking at Y as the dependent variable. That's right, but X is one of our covariates and what this is saying is if the missingness only depends on X for all Y, regardless of Y that's covariate dependent missingness. It only depends on covariates.

Question: And when you are talking about missing data are you talking about missing outcome data or-

DON HEDEKER: Yes, missing outcome data.

Question: Okay and not ****.

DON HEDEKER: Correct, missing the outcome, the Y variable, the dependent variable, severity of illness and what we're seeing with is that the probability of being missingness does not

depend of severity of illness in this part. It only depends on X variables by group like time, like group by time, so what this statement says is that you can have more missing data across time as you often do. You can have more missing data for the different groups and the amount of missing data per group can vary across time because those are all X variables, but it cannot, cannot, cannot be related to severity of illness. That is Y. This statement here is saying it's independent of Y. It can't depend on Y. Missing completely at random does not allow missingness to be related to Y, severity of illness. That is what GE assumes. It assumes that it can be related to X, time, group, group by time, not Y, not severity of illness.

Question: If it were related to severity of illness that would be **** or ****?

DON HEDEKER: The reason for it is this. It's kind of technical, but in GE, an estimation of GE it never forms the full likelihood for the vector Y. It doesn't have a multivariate representation of the dependent variable. It's treating each one. It's estimating this, this sort of separately, so there is sort of no glue among the values of Y. It's kind of technical, but GE in its estimation because it does things in a rather simpler way it carries this price. I can't explain it in a really good way without getting into the estimation details, but it carries this price in that it only allows the missingness to be related to X variables. Now the mixed model because they use full likelihood because they estimate things in a random or a complicated way, basically estimating the model parameters based on the multivariate distribution of the vector Y. It can allow conditioning on Y observed as well as X. Now this is an important step along also Y observed. Why is that? It can say missingness is related to what you've seen on the dependent variable. In other words if a subject is going along, not improving, not improving and the reason they're missing is because of that that's okay with missing at random. That is not okay with missing completely at random because the missingness is predicted by what you've observed on Y. In fact, in some studies they have protocols that say something like if a person's dependent variable like blood pressure gets below this level we're going to quit following them. In that case they're operationalizing MAR because they're saying that the missingness is related to their observed Y values.

The way to conceptualize as MAR, it's kind of like- it's sort of like you're taking the data you have on a subject and using that to impute what you don't have. You're not actually imputing. I want to make this clear. It's sort of the idea of that is what predicted, so if what you don't have is well predicted by what you do have, if you suspect that that's okay. That is with missing random. So missing at random and what we are doing with the mixed model goes just one step further. It allows the missingness to be related to X variables and also observed Y and in this case we call it ignorable non response, meaning we can ignore it. So everything that I've been doing is I've been ignoring it. I've operating under this MAR.

Now what-? But it says R is independent of Y missing. What is that all about? Missing not at random, so let me just say a little bit about Y missing. The way I've encountered this most is in this domain. I've done some analysis on smoking data where they're looking at perhaps maternal smoking across time. This is something that you might be able to relate to. In smoking studies when I began working in that area about 20 years ago almost they would often make this assumption, if they're missing they're smoking. They would say we conservatively recoded missing to be smoking. What they're saying is that the missingness is not missing at random,

that there is a perfect relationship between R, the missing indicator and what you have not observed, YM. They're saying the reason they're missing is because of the dependent variable that you have not observe. That is not missing at random. If what you would have observed is related to why the person is missing. Now one thing to realize is you can never confirm or refute that because it involves what you do not have. You don't have the missing observation, so you can't know whether they're related to the fact that they're missing. You can't test for it, but you can suspect it and in the smoking literature they suspect missing not at random. They suspect the reason the person is missing is because they're engaged in the bad behavior. You see this in drug studies. You see this is alcohol studies.

Now what I would argue about with that because I have several arguments about that. Number one, they say we conservatively recoded missing equals smoking. Which group typically has more missing data, the control group of the treatment group? The control group and now all those missing observations you're putting smoking, smoking, smoking, smoking. It's anticonservative vis-à-vis the treatment effect because you're putting in all of these bad responses for the control group. Here is another problem with that. You don't have that data, those data. You're putting in- You're imputing these data with a constant. What happens to your standard errors you think? They get small, so you're test statistics get overly powerful really. Again it's not a conservative assumption at all, so any kind of imputation using a constant like that, naive imputation rather not a good idea. Not a good idea.

So we've been analyzing under the missing at random assumption. Now I will say this. Missing completely at random and missing at random the difference involves what we do have, Y observed, so we can do tests between missing completely at random and missing at random and in my chapter 14 I describe some of the ways we can do those tests, but we cannot test between missing at random and missing not at random because again it involves YM. We don't have YM. If we had it we would use it, but we don't have it. That is the crux of the problem.

So missing not at random models, what you can do is this. You can simulate data under missing not at random and you can then estimate the model by missing at random models and see how badly it performs and I've done some work in that area and what you find is that they don't perform that well, so if the data are missing not at random, that is the reason the person is missing is what you would have observed and it's not related to what you have observed prior. That is an important point because again think about this. It's sort of the idea with missing not at random. Say a person is going down and then you don't observe this point. Missing not at random would be kind of they're over here. It's not well explained by what you have. If it's well explained by what you have that's missing at random, but if it's not explained by what you have and it's the reason that they're missing that is missing not at random and that is the problem. So if they are like that and again, you can never know that, but you can simulate that and you can show that standard models can yield pretty bad results. Now the important point is that the observed data give you no information about either to confirm or refute ignorability. You just can't do it. You can't test MAR versus missing not at random because it involves what you don't have.

What can you do then? Well you can estimate some missing not at random models and that is what I'll be showing you to get a feel and a sense for how robust are your conclusions across this

different class of models that make different assumptions about the missing data. That is kind of like the best you can do. So there is unfortunately not one golden answer per se. Now I will say this. In the clinical trial world basically what is being advocated, like what the FDA and organizations like that, they're advocating doing MAR analysis because that involves all of the data. That assumption doesn't involve what you don't have, what you can only speculate on and that is a less stringent assumption than the missing completely at random. So missing at random is kind of the gold standard at this point because again it's not as stringent as the missing completely at random assumption and going to missing not at random means that you've got to make guesses about some you don't have and FDA doesn't like that for good reasons.

Now what I'll talk about is two classes of missing not at random models and here the idea behind them are not really that hard, though the technical aspects can get a little challenging. A pattern mixture model, what you're going to do here is to use missing data pattern information in the longitudinal modeling, so for example, in this study here- Where we first developed this kind of model is we notice that 80% of the drug people completed, 67% of the placebo completed, something like that. We created a variable did they complete the study or not. We used that and interacted that with our model terms to see whether the results were consistent or not for completers, non-completers, so there we created just two patterns, did they complete the study or not and use that in our analysis. That was an example of pattern mixture model ****. So we're going to use information about whether they measured completely or not and use that as a variable in our analysis as a covariate.

The other class of models was called selection models. Here what you do is model both longitudinal and missingness concurrently and the longitudinal and the missingness are joined via shared parameters that allow you to say well here is the longitudinal modeling conditioning on the missingness model. I think in some ways selection models are a bit more complicated, though with NLMIXED we can actually estimate them within SAS, whereas before we had to use our own code to develop such models. Now I'm going to illustrate these models with mixed models, multilevel models, but I want to say that these classes can be done with structural equation models, can be done with GE models, really with all kinds of longitudinal models, so the pattern mixture selection ideas go beyond mixed multilevel models, but we'll illustrate those with mixed models.

Now again some comments on missing not at random. The ordinary mixed regression model and MR stands for mixed regression model and other full likelihood models assume missing at random, what is generally considered the gold standard. These extended models go further. Use on non-ignorable models can be helpful in conducting a sensitivity analysis to see how the conclusions might vary as a function of what is assumed about the missing data. Personally, it is not necessarily a good idea to rely on a single missing not at random model because the assumptions about the missing data are impossible to confirm or refute. They involve what you don't have unfortunately, so you got to use these models sensibly and to see. Again, the goal here is to see how robust our conclusions are as we explore these missing not at random models.

Let me describe a little bit first about pattern mixture models. Rod Little [ph], anybody know about Rod Little? Rod Little is not a little person with missing data. He is a giant in the world of missing data. He is little big man of missing data. He has written the book literally speaking, on

missing data. In fact, Little and Ruben it's a classic text on missing data. He has published several articles about missing data and pattern mixture models in particular. His '95 article in *Journal of American Statistical Association, JASA* is a classic. It's one of the great articles of our time I think in longitudinal modeling and I contributed a relatively minor article compared to those in *Psych Methods* with Robert Givens [ph] in 1997 describing applications of pattern mixture models.

How this came about was kind of interesting. I had been playing around with a schizophrenic dataset and noticing that there was more missingness among the placebo group and so I created this variable did they make it to week six or not and I put it in my model interacting with things and it was kind of interesting hopefully as you'll see, you'll agree that it's kind of interesting, but I didn't know what to call it and then I saw Rod Little give a talk about missing data and he described what are called the pattern mixture models and I realized what I have is an application of a pattern mixture model and so I wrote that up for *Psych Methods* as kind of like an illustration of pattern mixture models and that is where we got it published because the idea is not a difficult one. Rod Little he can think of difficult problems. I can't. I'm very pragmatic and so here what we did was we divided the subjects into groups depending on their missing data pattern and used that information then in our model as covariates because missing data pattern is a between subjects variable. It separates the different groups of subjects depending on how much data they have let's say.

For example, let me make this concreted. Let's say we have three time points, then there is eight possible patterns and here I'm using O to represent observed and to represent missing and again I'm talking about the dependent variable, whether it's missing or not at these different time points. Now there is eight patterns, but one of the patterns is really a challenging one, MMM. We have no information on them. Who are those people? Did you ever read about a study and they say well they identified a sample of say 1,000 subjects, but then they only measure 900 subjects? That is those 100. It's like they were intended to be measured, but for whatever reason they never got them at any time point. Now it's important to kind of consider those individuals. Are they systematically different than the people you have because when you run the model on the people you have you're not generalizing to the whole population, but a selected sub sample. So in Rod Little's work what he describes is how you can incorporate the MMMs into the analysis. Guess what you have to do, assume, assume, assume because you have no data on them basically, but you can assume various things about them and to see to what degree do your model conclusions change. That is the best you can do. Now again, I'm not going to consider- in the schizophrenic study I don't even know if there were such individuals, individuals that were intended to have been measured, but were never measured at all, so I'm only going to deal with people that have at least one time point.

Now how do we represent between subjects factors in our analysis? We can use dummy codes, so let's say we have these seven patterns. We can create six dummy codes and here what I've done is to characterize the six dummy codes using the reference group as the observed, observed, observed individuals and then D1 says well who is that group that is observed, observed missing different than the people that are completely measured, so these dummy codes represent deviations from pattern OOO. Other coding schemes can be used, so you don't have the necessary use these dummy codes. You can use effect codes, sequential codes. There is a bunch

of things that we can borrow from the **** literature really and then the idea is you put these variables in your model as covariates interacting them with your model terms to see how do your conclusions depend on missing data pattern. That is the essential idea.

Now admittedly six dummy codes is going to create a lot of model terms, so what you might do is to collapse those groups, so the first thing you want to do is explore how many missing patterns do you have, missing data patterns you have and if don't have too many or you don't have to many missing data you create less patterns recoding some together, things like that, kind of like what we do often in survey data when you've got a variable like for example you might have marital status and you've got seven different categories to marital status, but then in the actual analysis you find out well I've got to put these two groups together because I only have three individuals like this and ten like that, same kind of idea.

Now here in our paper in '97 we did a very simple example of pattern mixture. We just said did they make it at week six or not, so just two patterns basically. We're putting a lot of people together in the people who are not measured at week six. Did they make it to the end or not and again, as I've seen in the placebo group there is much more dropout than in the drug group and if you do a test that is a significant result, so it's true that dropout does vary by treatment group, but here is what we want to explore with pattern mixture. Is the dropout related to the dependent variable and maybe more interestingly, does dropout moderate or change to influence of other variable's effects on the dependent variable? Does it change our conclusions?

Now it always worried me that there was so much more dropout in placebo than drug group and then I was at my dentist reading the *New Yorker Magazine* and I found this slide ****, member placebo group, so then I went to my lawyer and said when I die on my tombstone I want to put analyst, placebo and drug groups, as if I have a lawyer or a tombstone, but yeah, that was kind of curious to see that.

So schizophrenic data, let's look at this with pattern mixture modeling. This is what we've done so far. This is a simple random intercept and time model. What we're going to add to this is a new variable, dropout. Did they get measured at the final time point or not and I'm going to add that and interact that with my other model terms, so I've got dropout. I've got dropout by drug, dropout by time, dropout by drug by time. So in this way of writing the model the parameters B **** to B3 are for the completer sub sample because I've defined dropout to be one if they dropped out of the study, zero if they didn't, so these parameters related to the completers, the people who made it week six and the remaining ones are how are dropouts different from the completers.

Now in this analysis I think at least, the three-way interaction is of particular interest because that indicates how the drug by time interaction varies depending on study completion. So just like in this analysis the thing of most interest was beta three, is there an interaction of drug by time, do the groups have differential rates of change. Now I think the thing of most interest is this sort of beta 3D and here I've just used these Ds as superscripts to indicate that these are basically the same parameters as their cousin you might say, but just dropout different from the completers.

Now this picture tells a lot of stories, a very useful picture. This is a means across time for the completers, two groups; placebo and drug and dropouts, two groups; placebo and drug and what do you see when you look at this in relation to the quote, drug effect? The drug effect is the divergence really of the two lines. It's bigger here. It goes from something like this to like that. When I just plot these means down I was like struck because to be honest I was a little bit worried that if I go into this I'm going to lose my drug by time interaction and then the drug company won't pay me. No, I was just kidding. No, but I was a little thinking like maybe this will wipe away my drug effect, but when I looked at this plot I realized if anything it's going to increase it.

Question: So in certain ways you say that the placebo effect was not experienced by the dropouts.

DON HEDEKER: Placebo effect, what do you mean by that?

Question: Like a the sense that psychologically you receive that pill, so you feel as though you were treated and so you then report better-

DON HEDEKER: So are you saying that because this line seems to go down a bit versus this one being more horizontal?

Question: Yeah.

DON HEDEKER: Yeah, that could be.

Question: The psychological **** dropout.

DON HEDEKER: That they're not experiencing the placebo effect.

Question: There should be like a non measurable ****.

DON HEDEKER: Yeah, you're right. I mean it could be. It's true that when I look at this line it's rather horizontal compared to this counterpart, which is a bit decreasing, so it does seem like that among the placebo group the ones who are dropping out are not doing too well. What I was thinking was kind of like well maybe the psychiatrist was monitoring them and pulled them from the study. I don't know this. I'm just wondering. Why among the placebo group-? People who seem to be dropping out are pretty flat on average. They're not really improving, but it could be exactly what you're saying. We don't- Yeah, it's hard to know. The main thing I think though is again this divergence of the lines looks here more pronounced than here. So that picture really is the story. Yes.

Question: I don't know what- I may stand corrected, so the missing, you have a missing dependence there, so if you include the dropout as a covariate how do you code the dependent ****?

DON HEDEKER: Right, it’s not- It’s just missing. It’s not there. These people are only measured up to here. That’s it. I'm not imputing anything. I'm just using the data I have for this. For these subjects here they only have data at time points up until this one. They don’t have data at time point six. These subjects have data at time point six, so all I'm doing is I'm creating a variable that says were they measured at time point six or not and I'm putting that in my analysis as a covariate interacting it with my other model terms, but I'm doing no recording or imputation or anything for the missing dependent variable.

Question:**** is missing **** among ****.

DON HEDEKER: They’re included, yeah.

Question: They’re included.

DON HEDEKER: Yeah, they’re D subjects. Think of it this way. Dropout, completed or dropout- It’s just like- Say it’s gender. It’s could be any grouping variable. That only thing that sort of distinguishes it is that it’s kind of related to the missing data pattern, but it’s just like a subject variable. That is all it is and again, my model doesn’t require that everybody has every time point. It doesn’t require that, so these people are in as they are just like these people are in as they are. They have different amounts of dependent variable information. That’s true, but that’s not a problem.

Question: So the subject is missing. The subject should have at least one, one time point.

DON HEDEKER: Yes, that’s correct.

Question:**** the study how the drugs were administered? In other words, did they have to go to the site to **** or did they have to go to the site to ****? **** go to the site prior to this? That would easily explain that if the ones- if you pull the ones the dropped at and those are the ones that have the most ****, which are the most disorganized folks and least likely to show up, which would strengthen the drug and drop off the placebo effect.

DON HEDEKER: Yeah, I don’t know. Unfortunately, I got the data. I don’t know, but you’re right. That could be it because you’re right. The other thing that is kind of interesting is not only are the dropouts, placebo rather horizontal. Conversely, the dropout drug patients they’re improving.

Question:**** inpatient.

DON HEDEKER: I think they were. The **** dataset was inpatient. Yeah, this one I don’t really know, but yeah, it seems like again based on these two plots the first thing to glean is that okay this goes to that. Then it does seem like maybe the placebo dropouts are worse than their completer counterparts. Conversely, the drug dropouts are better than their completer counterparts. Now we’re actually going to assess those things, but the picture really is pretty compelling I think for these data.

Now when we did the book I kept wondering, missing data pattern. I've only got two patterns, did they make it to week six or not. Was I being too simplistic with that? And I thought well now is my chance to correct my errors or my ways and so for the book we went into a less restrictive pattern mixture model. I used week of dropout in forming the missing data patterns. In other words, there are some subjects that made it- they were measured at week zero. They were measured at week one. Thereafter we didn't see them. They were measured at week zero. They made it to week three. Thereafter we didn't see them.

Now in all of the analysis I've just described those are all described as dropouts. They're all put together in one pattern. Now I'm going to disentangle them to find out does it matter when they dropped out of the study. So here I've got six groups, five drop out weeks and completers. In other words, people dropped out at week one. Then it was they were measured at week one, but not thereafter, week two, week three, week four, week five. So previously what I was calling dropout I'm not going to call five different dropout groups depending when they were gone.

I'm still going to use the completers as my reference cell and so I've here a more extended model in which I still have as completer parameters beta **** through beta three, but then I additionally add these indicators to how dropout group M, and there is five dropout groups, differ from the completers in terms of the intercept, in terms of the drug effect, the time effect of drug by time interaction, so there is 20 beta parameters here, a lot of parameters. I mean that is why initially I didn't go to this. I thought this was a little much, but then I thought in a book you can put in anything. There is no reviewers. Yeah, that is the problem with a journal article, those **** reviewers. You're doing a book you don't have to worry so much about that, so here I went a little bit hog wild. Now I'll admit when I went hog wild I was a little worried. I was thinking like again that article in '97 we're going to find something that we glossed over.

Now in this slide what I'm showing are just the beta estimates from the three models. The first on the left-hand side is the model that ignores dropout. It allows incomplete data across time, but doesn't get into this pattern mixture modeling. We've seen those results before. That is exactly what we got with PROC Mixed for example, no drug effect at times zero, effect of placebo, drug by time interaction. The next, the middle, this is the model, the pattern mixture model of 1997 when I said I put all of the people that dropped out regardless of when they dropped out together as a group, as a pattern and I contrasted them to the completers and then on the right-hand side here- I guess it's- Let's see. Well it's your left- You're your right-hand side. Stage left or something, I don't know how that works. There I've got all of these different dropout patterns compared to the completers and notice that some of them don't have many observations. I mean this one has five observations and I'm trying to estimate four parameters, a bit much it would seem. I was kind of curious.

First thing to take a look at is this. I'll get into the estimates in a moment. Let me just- Sorry, I don't- We can do likelihood ratio tests comparing these. First of all if you do a likelihood ratio test comparing the simple pattern mixture model to the MAR model that is a difference of chi-square of 26 on 4 degrees of freedom. That is significant. What that is saying is that this model here fits statistically better than this model. If you compare the hog wild pattern mixture model to the MAR model again that is 42 on 20 degrees of freedom. That is also significant. However, if you compare the hog wild pattern mixture model to the simple pattern mixture model that

difference is about 16 on 12 degrees of freedom. That is not significant. So what that means is that simple model that we described in our ’97 paper was really far enough. We didn’t have to go hog wild as this model. So combining all of those dropout patterns into one group was a rather sensible thing. I didn’t know that actually when we ran this model because I didn’t do this and I was kind of reassured to see that when we went to this more complicated model.

Now let’s take a look at those results for the middle, the pattern mixture model. The first four estimates up there are estimates for the completers. I've got the estimates, the standard errors and the P values, so among the completers, that is among the completers, the drug effect. The drug difference at time zero is estimated to be .2, about a fifth of a point and sort of marginally significant. What that is saying, among the completer sub sample the placebo are a little bit more severely- Sorry. The drug are more severely than the placebo counterparts at week zero and that is not a significant result, but in that direction.

The time effect, that is the time effects for placebo group, that is significant, negative. The drug by time interaction that is negative and significant. In fact, if you look at those estimates there they are essentially precisely what I've previously put in this slide here as the estimates from the analysis of completers only because that is basically what they are. They are representing the model parameters for only the people that made it to week six, so our estimates there or our conclusions are really what we saw here on slide six.

Now what we additional have though on slide l7 is how are the dropout patients, the hundred or so people that didn’t make it to week six, how are the different from the people who made it to week six? The big one that comes out is the three-way interaction, .002, negative. What does that mean? It went from this to that. The divergence between placebo and drug was even more pronounced among the dropouts than among the completers. It’s significant for both, so this is a good result in some ways. The drug by time interaction is very robust. It is there for the completers. It’s even more there among the dropouts. So doing this analysis has in some ways, given us more confidence about the drug by time interaction.

Now what you can do then is to take the estimates and then again, see how well they fit the observed quantities for completers and dropouts and they fit rather well and the main story here is that divergence of the lines, the drug by time interaction greater among the dropouts than the completers, so if anything there is bigger effect among the dropouts.

Now we sent in that paper at that point for the first version of it saying how there is this interaction of dropout by drug, by time basically interpreting that saying that the drug by time interaction was more pronounced among the drug group. Sorry, dropout than among the completers and the reviewer was very astute and said, “There is a final step in the pattern mixture- “

Don’t you hate when that happens?

Question:****.

DON HEDEKER: **** eject button, right. Good bye. So we sent in the paper. We got it back. They said there was one- one reviewer said, “There was one step in the pattern mixture models that you’ve neglected to do.” “You’re paper is rejected.” “We invite a resubmission.” So that is the typical thing you get when you send in a paper. You never get it accepted the first time around and I'm glad that they did reject it because there was a step that I had failed to do because I didn’t know anything about it. I didn’t read the literature very well. So Rod Little in his ’95 article described about how you can get average results across the missing data patterns. Why do you want to do this? Let’s say you go to the FDA and you say well among the completers there is a significant drug by time effect, but among the dropouts it’s even larger and that is what we found. They’re going to come back to you and say, “What is the drug by time effect?” You say well among the completers it’s this. Among the dropouts it’s this. What is the drug by time interaction? They don’t want a stratified answer. They want one answer, so in the final step what you can do is to create that one answer going over these stratified missing data patterns, so obtaining average estimates over the missing data patterns is kind of the final step in the pattern mixture analysis, the missing data patterns. For example, here what I'm considering is completer and dropout.

Now to do this we can use sample proportions as estimates of our missing data pattern proportions. Those estimates do depend on quote, our model from missing data pattern. What do I mean by that? If we look marginally completers and dropouts there was this many completers, this many dropouts, so if use these weights for completers and dropouts my model for missing data pattern is really no model at all. It’s just using the marginal proportions, but I already know that completer, dropout status varies by treatment group, so instead I can use these weights for the placebo for the placebo parameters, these weights for the drug parameters and that would be perhaps a more sensible approach. To be honest it’s not mattering much for this particular dataset whether I go with the simpler marginal proportions or I go with the group stratified proportions, but it seems like the latter a bit more sensible here.

Let me show you what I mean. What we’re doing is this. We’re getting an averaged estimate averaging across the completers and the dropouts weighting each by the proportion of completers, the proportion of dropouts. They’re just weighted averages. That is all they are. Now when I looked at that equation I realized in my model what I had is not necessarily the estimates of the dropouts, but I had the estimate for how the dropouts are different from the completers, what I'm labeling as beta **** D and I played with that equation a little bit and I realized well the average estimates in that case are just equal to the estimates for the completers, plus how are dropouts different from completers weighted by the proportion of dropouts that there are. So that is all that I have to do. It’s simple arithmetic. It’s just taking the set of estimates for the completers, the set of estimates for the dropouts and combining them in this simple way.

So the beta had C co-response and the coefficients in the current formulation not involving dropout. That is beta ****, beta one, beta two, beta three. Beta **** D corresponds to the dropout **** coefficients in the current model formulation, dropout, dropout by drug, dropout by time, dropout by drug by time and Pi **** D is the sample proportion of dropouts. Again for this I can simply use- Sorry, this sample proportion of dropouts. If I want to be a little bit more sensible I can use these for the placebo and these for the drug.

So here is how that arithmetic goes and incidentally, what you see here is basically these average estimates are combinations of the model estimates, so in SAS what we can do is to actually get these, obtain these using estimate statements in our analysis, so I'll show the arithmetic, but then I'll show how you can actually do this with SAS. Here is the arithmetic and this is using the marginal proportions as the weights. So there is 335 completers. There is 102 dropouts and what I have in parentheses are the parameter estimates from my simple pattern mixture analysis. Beta **** was estimated to be 5.22. That is the intercept for placebo completers. Placebo dropouts differed from that by this .32. So I can either do the calculation that way or do the calculation this way and come up with the estimate. The average estimate across these two groups as 5.3 and you can see it more towards this than this because there is three times as many completers as dropouts. It’s just a weighted average.

The time effect, similarly I can do the same kind of arithmetic. The drug difference at time zero, the drug by time interaction. I do this bit of arithmetic weighting the completer estimate, the dropout estimate. Getting their weighted average I get these four estimates. This is what the FDA wants. They don’t want the stratified results. They want the combined results.

Now here is how you can do that in SAS and incidentally, this program is included among one of the SAS syntax files that you have. So pattern mixture, random intercept and trend model using marginal dropout proportion to estimate averaged results. All of this should look pretty familiar at this point. The additional terms that we’re putting in here is this variable for dropout, did they make it to week six or not, dropout by time, dropout by drug, dropout by drug by time. This analysis right here produces precisely that set of result estimates that I had in the middle of that previous table, so that syntax produces precisely the estimates I have here. This is what I'm doing right there.

Now to get the average results across the two missing data patterns I can use estimate and what I'm doing with the first one for example, I'm saying let’s get the average intercept. Well the average intercept is going to be equal to the intercept for completers plus how are dropouts different from completers in terms of the intercept multiplied by the proportion of dropouts that there are in my sample. In fact, I could make this estimate statement easier by- you notice I've put everything else with zeros. If you don’t include them in the estimate statements SAS assumes that those are zeros, so this business and all of this business could be removed and you’d get the same result, but I've sort of put it just for conceptual reasons. You can see exactly explicitly what is being done.

The average week effect, the week effect for the completers plus the proportion of dropouts times how are dropouts different in terms of the week effect. I'm explicitly doing exactly the averaging that was described on the previous slide, same for drug effect, same for drug by time interaction. And this .2334, what is that? Well that is the proportion of dropouts in the sample. So using estimate we can get these averaged results. I didn’t know this back in 1995 or 6 when I was working on this paper, but I know it now.

Let’s go over these last- these few slides. This is more if you’re using SPSS and can’t the estimate statement. There is other ways you can get these averaged results, but I think I'll skip

over these in lieu of time and it doesn't seem like many people use SPSS here anyway. Or do they? So let's skip on to page 27, so the other slides they were all about how to do this within SPSS where you don't have the estimate statement.

Now if I run that SAS procedure using the estimate statement SAS will print out the estimates for these four averaged parameters, intercept, time, drug, drug by time and it print out standard errors for them and give P values. Great, we're done, right. What do you think? No, unfortunately, we're not quite there. We're almost there, almost there, very close, very close. I got some good news. I got some not so good news. The averaged estimates obtained in this manner are fine. What I mean by that is this column here absolutely fine. What is not so fine is this column, the standard errors and as a result of that what is also not so fine are these P values, so you might care little about the standard errors, but you probably care about the P values, so we have a little bit more work to do.

Standard errors are not exactly correct. Why is that? Well let me go back to- Sorry. I'm sorry if I'm jumping around, but let me just go back to this slide. What I've done is to use this equation to get my averaged estimates and in SAS I've used the estimate statement and I've put this sample proportion in, .233 or whatever. SAS thinks that that is a known constant, but it's not. It's an estimated quantity. In other words, there is some uncertainty about it. It's our sample estimate of how many dropouts there are. It's a sample based- there is some uncertainty around that. I have not taken that uncertainty into account. What does that mean? The standard errors that I've gotten thus far are a little too small because they've taken into account the fact that these are estimated parameters, these are estimated parameters, but not that that sample proportion of dropouts that that's an estimated quantity as well. It has not taken that into account. So I have to take that into account at least to get published in *Psyche Methods*, so here is how you can do it. So great, we're done. Not exactly.

The standard errors for the average estimates they consist of the piece that I have considered thus far. I'm sorry about this dense notation, but the first part there on the right-hand side of the equation is the **** treating the sample proportion as known. In other words, the square of the standard error using this formula and not taking into fact that this is an estimated quantity. So this is what SAS is currently printing out when it prints out those standard errors, but there is an additional part involving the uncertainty in the estimate of how many dropouts there are. That is what I need to consider. Now if you look at that equation long enough you'll probably decide it's about time to go, but one thing to realize, this is a rather small contribution because in the denominator here we have the sample size to the cubed power. Say that is 400 and blah, blah, blah cubed. That gets to be a rather big quantity that you're dividing things by, so this augmentation that we need to do to be precise is of little practical value to be honest.

Here shows a little work that I did to get the right results. These are the averaged estimates I get using the estimate statement. In parentheses here are the standard errors that it prints out. These are standard errors acting like that proportion is a known constant. I'm squaring them here to put them on variance metric. I'm then dealing with that right-hand side, the uncertainty in using the sample proportion in this business and that leads to what I call the augmentation of the variances. Now if you look at this augmentation these are all rather small. These are added to this to yield the correct variances for these estimates. The square root of these give the proper standard

errors. So what is the difference between this column and this column? Not much. Why? Because the sample size is so large. It doesn't matter. It's one of these things that statistically I have to do it, otherwise the police are going to arrest me or something like that, but it makes very little practical impact, so you can really just use the estimate statement in SAS to do this kind of work. Doing this additional augmentation is of little practical gain. Yes, question.

Question: I was wondering how **** the standard errors are **** because you're taking it to be a constant. However-

DON HEDEKER: They're not inflated. They're deflated.

Question: Okay, but you're saying, but we also looked at the fact that dropout was not related to time ****, so **** not as ****.

DON HEDEKER: Are you talking about when I say not related to time, this latter column?

Question: Right.

DON HEDEKER: But I am- Okay, I do have dropout and I'm using the sample proportion of dropout patients. I'm using that as a known constant, but it's not known. It's just based on my sample. In other words, it's an estimate and when I use it in the estimate statement in SAS that is acting like it's a known value. It's a sample estimated quantity, 100 out of 400. There is some uncertainty attached with that. It's just like if you estimate who is going to win the presidential election. If you base it on 400 there is some uncertainty around that estimate. If you base it on 1,000 there is less. It's more like a constant. I'm basing it on what is the proportion of dropouts in the population based on this sample of 100 over 400 or so. So there is some uncertainty with that that I am not acknowledging when I do this. Here in this it's acting like these are known numbers. They are not known numbers. They are estimated numbers. There is some uncertainty attached to them. There is some variation attached to them. It's kind of a minor point pragmatically speaking as I found out, but it is a point.

Well let me in the time we have left then I'm going to skip over the- Let me go to this. This show then slide 20- 33 where we've been thus far we had the ordinary analysis not dealing with dropout. We went to the pattern mixture analysis. We saw the significant relay interaction. Now what we've just done is a fair amount of work to get average results, pattern mixture average results. Now if you look at these and you look at these you're going to say was all this work worth it and I would say yes, I got a paper out of this work. For me it was worth it. No, it's really it is reassuring. What it is saying is that the drug by time effect is robust across these two different approaches to missing data. This approach is essentially ignoring the missing data. It's acting like the missing data are missing at random. This approach is saying well they might not be missing at random. They might be different for people who complete and incomplete and then let's average across those, but the results are rather similar. So I've got good confidence in my drug by time interaction by doing this pattern mixture modeling.

Well in the final time we have left let me just talk a little bit about selection models because again it's another class of models for longitudinal missing data. Now these models have been

called a few different terms. Rod Little called these random coefficient selection models. Hogan and Laird [ph] called these random effects dependent models. Carol Hayes [ph] called them shared parameter models. You might be saying why can't you agree? Why can't you just call them one thing? Let me ask you this. In the **** literature do people call things the same way all the time? No, so we have our differences too. So this is another class or interesting models for missing. What you do here is you specify a model for the outcome and a model for dropout and they share the random effects and what that means is you get estimates for the longitudinal model conditioning on the model for dropout or adjusting for the model of dropout or controlling for the model of dropout. That is the notion here.

So what I'm going to use for my longitudinal model is exactly the same mixed model I've been using, but for the dropout model what I'm going use is the survival analysis model, modeling time until the subjects dropout. Now in this particular case time to dropout could only be one to six because there is only that many time points. In usual survival analysis you have time until the event as a continuous kind of variable, days until the heart attack or something like this. You guys don't have days until the heart attack though. But you'd have a continuous variable of time. Here I have really a finite or discrete measurement of time until the event, so I'm going to be using a special class of survival analysis models called group or discrete time survival analysis models. DI is the time of the dropout for subject I, did the dropout at week one, week two, week three, week four, week five and I'm building here a survival analysis model for time until dropout. Now in this survival analysis model I'm going to have some variables and most importantly I'm going to have the random effects from the longitudinal model. That is what is shared between them. That is sort of the glue that allows us to therefore estimate the longitudinal model parameters controlling or adjusting for time until dropout.

So two models, but they share the random effects and therefore the longitudinal model estimates, the betas are adjusted for time until dropout and to the extent that these alpha, star parameters are non zero, that is the random effects have influence on time until dropout. This is classified as a non normal model because missing this here characterizes time until dropout is dependent on observed and missing data via the random effects, so that I kind of a technical point, but the issue is that again we're not assuming MAR per se by having this glue between the two models.

Now here is my time until dropout model for the two treatment groups, placebo and drug. So you can see that few people only made it to week one and what is meant by this is that they were measured at week one. They're measured at week zero. Everybody is measured at week zero, at baseline and then at week one 13 placebo patients were measured, but we didn't see them after that. 24 drug patients were measured at week one, didn't see them after that. So max week here is what is the last time point that we saw the subject, that we got data on and you can see that here dropout is more common among the placebo group and in fact, if you do a Pearson **** score test of this it's a significant result and if you do a test for trend it's even more significant. Why, placebo subjects tending to dropout earlier relative to drug patients. The drug patients are tending to stick around.

So here is my two models that we're using. Longitudinal model is exactly the same thing. It's the random intercept in trend model. My dropout model is going to be a time to dropout. This business here is a little bit technical, but what I'm doing is to standardize the random effects, so

that they're on the same metric, so they have zero mean and unit variance. So I won't get into this in detail, but all I'm doing here is to re-expressing the random intercept and trend as a random intercept and trend now that are on the same metric. It's like standardization in regression where you get standardized estimates. That is all I'm doing. The variables now have zero mean and variance one. Why am I doing that? I always try to make things more complicated than they should be. No, I was kind of interested in knowing is it more about their starting point or their trend that is related to dropout and I thought if I want to make that comparison then I should have the variables on the same metric, otherwise my estimates they're going to be on different metrics, so I was just trying to put them on a common metric. That was the motivation.

My dropout model is this survival model. Incidentally, what that uses is what called the complimenting log, log link. The reason for that is for discrete time survival data that gives the proportional hazards model estimates, so it's like I'm estimating a Cox [ph] model, but a specialized Cox model that can handle the fact that people only have these finite number of dropout times because in general in survival analysis people can dropout or have the event over a wide range of values. Here there is a rather finite number of times they could dropout. There is a lot of ties, what is called in the survival literature a lot of ties.

So in this model I'm putting in the effect of treatment group, the two random effects intercept and the time effect and then interactions of drug group by the intercept, drug group by the time effect. This shared parameter model is **** normal to the extent that the alphas are not equal to zero. Again, this is a test of whether a particular model's ignorability is reasonable via a particular model of non-ignorability. It's not a general test of ignorability. We don't have the data by which to do such a global test.

Here is the estimates I get. Now I've done all this actually in SAS and I was quite astounded that I could do this in SAS because this a- you know it's a bi-variant model where I've got longitudinal data I'm modeling, time until dropout I'm modeling and I'm modeling them at the same time, so there is no like PROC longitudinal survival available. That doesn't exist, but NLMIXED because it has programming facilities I could figure out a way to do it. Now the coding is not super simple, but it can be done. I'm not going to go into that in great detail, but if you're interested in that it's attached as one of the files. This is the result of what I found.

Now the first thing I did was to estimate the model with that code where the two were separate. The longitudinal outcome was separate from the time until dropout. In other words, the random effects from the longitudinal model did not influence time until dropout. Now why did I do that? The reason I did that is that if I did it that way I should get exactly the same results for those two models as if I ran those two models separately, so I wanted to confirm that my code was correct and it was. So those estimates up here are exactly what we've seen all along for the mixed model for the longitudinal outcome. This is what I got from the survival analysis using only drug as a predictor of survival. It's negative. It's protective. There is less dropout in the drug group. It's a protective effect for time until dropout. It's a significant effect.

So the separate parameter model gives identical results as running these two models separately. Then I knew my code was working correctly. Then I did the shared parameter model adding in

the two random effects, **** slope and their interactions with drug group. Now if you look at this it's kind of interesting. First of all the random intercept effects are not really significant per se. So in other words where a person starts out in this study is not predictive of time until dropout, so it's not how sick or well they are in the beginning that predicts when they dropout, but the time effects are predictive of time until dropout and not only that. They're related in an opposite manner.

Let me explain. In this model the random slope effect that represents the effect, the time trend of severity of illness for the placebo group and that is positively related to dropout. Placebo subjects are dropping out quicker. That is what that is all about. It's .06. You might quibble with that, but that is the direction of the effect. Placebo groups that are having a higher trend, higher trend are more likely to dropout, the ones who are not improving. Conversely, for the drug it's negative and highly significant. So what you're seeing is that opposite reasons for dropout. Among the drug group it's the people who had the more negative slopes that are dropping out. In a way this confirms exactly what we saw in the pattern mixture model. The placebo group subjects it's the people who are more horizontal that are dropping out. For the drug patients it's the ones who are more negative that are dropping out.

So that is that part of the thing. Now perhaps a more interesting thing is the result of the longitudinal parameters here. Now again, this model here is under MAR, missing at random. This model goes beyond. Do our conclusions change? Well the drug difference negligible, so drug at times zero no different. The time effect, that is the time effect for the placebo group. It's significant in both, but notice that it's a bit less than what it is here, so again, adjusting for missingness here has said well placebo effect there is one, but it's not quite as great as if you ignore the missingness. The drug by time interaction conversely, it's a bit greater. It's significant in both, so again the conclusion is the same, but if anything, just like what we saw in the pattern mixture analysis, if anything, looking into missing data we have more confidence in our drug by time interaction. If anything it's greater. It's not lessened.

So for longitudinal component the conclusions are really the same as the MAR, but perhaps strengthened. In terms of the dropout part again we saw a marginally significant slope for the placebo group. There is a tendency to dropout as the slope increases. This is a slope for severity of illness across time. Significant drug by slope, the slope effect is opposite for the drug group. Drug patients with more negative slopes, i.e., greater improvement more likely to dropout.

So let me present to you this final slide of results and when you see this you're going to all want to come up here and beat the crap out of me for all the misery I put you through in the last couple of hours. Why do I say that? Here we got results for these model parameters, for completers, for all cases. Shared parameter we've just gotten. The pattern mixture average results, all this work, all this work. Are our conclusions any different for any parameter? No. Is that bad news? Not really. It's good news. At first I figured that's bad news because like you want to showcase like hey if you do it this way you find this. Well in terms of these we're seeing the same pattern of results. That is why I say you're going to beat the crap out of me. But look at what we've learned. No, I'm serious. I'm not being facetious.

When we did this analysis, when we started off this afternoon's session we found out that there is a time effect, placebo response. There is drug by time interaction. Since that point what we've discovered if anything, among the people who dropout there is a bigger drug by time interaction. That is what the pattern mixture model told us, so among completers it was like this. Among dropouts it was like that. Now the shared parameter model what that **** told us is who is likely to dropout and what we saw is that reasons for dropout were kind of different for placebo and drug group. For placebo patients it was the people that were not improving that much that had higher slopes that were more likely to dropout. For the drug patients it's conversely, the people who had more of the negative slopes that were likely to dropout. Now all of that we didn't know when we just did this.

Now admittedly those things have not changed our conclusions with regard to these four parameters, but I think in investigating those things we've gotten a better handle on the missing data and why people are missing and how that is affecting our conclusions.

So having said all that let me conclude- a few concluding remarks and then we'll convene for today and yesterday as well and you'll be free to go and be rid of me. I would say hopefully these last few days you've seen that the mixed or multilevel models they're useful for clustered as well as longitudinal data where you don't have the same number of observations per cluster or in longitudinal you don't have complete data across time.

Now these models can handle subjects measured incompletely or at different time points as we've seen. We assumed that the missing data are missing at random, that is they can depend on covariates and available data on the dependent variable. That is the MAR analysis that we've been conducting, but we've also gone beyond that to look at these pattern mixture and selection models. They augment the mixed model. Pattern mixture I think a little easier to entertain. It adds missing data pattern as a subject variable and you then put it in to the model, interact it with your model terms and you see to what extent is missingness and here, the reason I'm putting missingness in quotes is it depends on how you form these patterns. So I'm using it kind of in quotes. It depends on how you're defining that and the way I defined it in my main analysis was whether or not they made it to the end of the trial or not. Does that influence the outcome? Does that interact with model terms, group, group by time interaction? And we saw that it did interact with group by time interaction in our example.

Selection models go a slightly different route. They look at missingness in terms of important covariates, missingness, time until dropout in terms of the random subject effects and they can therefore give you estimates of longitudinal adjusting or controlling for the missingness process.

Now I want to stress this, this bottom line. In all of this I have not imputed any data. I don't mean that to be a knock against imputation because that is clearly a useful tool in many contexts, but I have not used that here. I just want to clearly define what I have done and what I have not done. I have not imputed any data. What I've done instead is try to maximize the information I have out of the incomplete data and to see to what extent being incomplete or not makes an imprint on my dependent variable.

So that is about it. This shows the God awful code to do the pattern mixture models and subsequently the shared parameter models, but I think we'll end at this point. You can look at that code again tonight when you're having trouble sleeping and if you haven't done so already, please fill out the evaluation form and if you have any questions I'll stick around and be happy to answer anything that you have question with and again I want to thank you for your attention these past two days, over two grueling days. Thank you.