Multilevel models for longitudinal continuous outcomes, part 1—2010 MCH EPI pre-conference training workshop

DON HEDEKER: So and then we're going to break for lunch at noon, so and then we'll conform to the schedule as given, but before we get to that any questions about what we've gone through thus far this morning?

One question that came up was an interesting one about a study in which the subject had IQs of individuals across time and then the idea is well what do you put in as time. For example, you have their age. If you've got age, age would be-what would be considered time presumably if you're interested in how does a person's IQ or in this case a baby's IQ change as they age and there clearly the age variable would be quite different for the different subjects in your study because you might have them measured all at birth let's say and then some at three months, some at four months or whatever and so age would be like what is time in my model, but again, it would be allowable to have subjects with different ages across the many different time points that they're measured.

Well let's get into the analysis that looks at diagnosis that is group, endogenous versus non-endogenous. Now in this study group diagnosis has only got an I subscript. What that means, it doesn't change across time, so a subject is diagnosed as either being endogenous or non-endogenous and within the depression episode they stay in that across the six weeks, but what I'm looking at additionally and really the thing of most interest here is the group by time interaction because the diagnosis by time interaction gets at do these two groups have the same rate of change across time or do they vary in the rate of change across time. In clinical trials, once you're doing clinical trials the group by time interaction is usually of most critical interest because what you're interested in finding out is do people that are randomized to some kind of intervention do they change across time in a better way than the control or placebo group. So group by time interaction is really of most interest. Here again, it's going to get at do the endogenous, non-endogenous patients have different rates of change.

So in terms of the regression model I have this model and this is very similar to the slide I had earlier where I had males and females as two different groups that had different intercepts, different slopes and I had individual variation both in terms of the intercept and in terms of the slope.

So on the top part I've got the mean model, which says that basically we've got an intercept, an effect of time, a group effect and a group by time interaction and then at the subject level I have subjects deviating relative to their group in terms of their intercept, deviating in terms of their group relative to their slope. Again, here I'm coding time in a real simple way, just zero to five

for the six weeks of the time points and I'm using group in a simple way too, zero for non-endogenous, one for endogenous.

Now here is the multilevel way of writing the model and what we see here again in level one is exactly the same model as we have considered up until this point. We're saying that the Hamilton Depression Scale is affected by time, but now at level two we're saying a person's intercept varies depending on what group they're in plus their own individual deviation to that.

So Beta not [ph] let's think about and interpret these parameters. Within subjects model B, zero, I represents the week zero Hamilton level for patient I. A person's week zero Hamilton level depends on what group they're in and their deviation to that. Diagnosis is coded zero for non-endogenous, one for endogenous therefore beta not represents week zero mean for non-endogenous patients. Beta two represents the difference in week zero means for the endogenous patients relative to the non-endogenous patients and epsilon zero I is how is a person's week zero Hamilton level different than their group mean, so beta not and beta two are about the group means when time equals zero and epsilon zero is how is a person different from their group mean at time equals zero.

Analogously the model for the weekly change for a patient is modeled in a similar way. Beta one represents the slope when diagnosis equals zero. That is for the non-endogenous group. Beta three represents how is the slope different for the endogenous patients relative to the non-endogenous. Beta three is the group by time interaction, are the rates of change the same of different for the two groups and epsilon one is a person's slope relative to their group mean. So this has got the words for these different things.

Now when you write it in this multilevel way it kind of indicates something that you would not have realized perhaps from writing it in this matrix way. What is that? Let's consider this model and relative to the model we fit earlier this morning. The model we fit earlier this morning didn't have these two. That is what we considered. We previously did not consider yet group. Now we're considering it. In a linear regression model what happens to the error variance as you add in important repressors, independent variables? It goes down, right, so think about this model first here. Relative to what we've fit now we're adding in this. If diagnosis explains why people are different at week zero there is going to be less subject variation in the intercept because we'll have explained some of it by knowing what group they're in.

So let's go back to the estimate we had in that previous model, so let me go back to slide 17 and I'll jot this down. The intercept variance in that model was 12.63 and while we're at it let's jot down the slope variance as well. It was 2.08. Because what I'm arguing here is if beta two is a significant effect then there is going to be less individual variation in this model because we'll have explained some of it. Likewise, if beta three, if group by time interaction is an important predictor they'll be less slope variation. It kind of comes back to the question yesterday when we had clustering and we can have the clustering variance and then we might look at cluster variables and try to explain some of that variance. I mean that is what I'm doing here is I'm adding in variables to try to explain why people are different at time zero, why people are different in terms of their slope and the ability to which I do this that will potentially diminish the individual intercept, the individual slope.

So again, that aspect I don't think you would glean that from writing it this way or from just going to PROC Mixed and running things. I certainly didn't appreciate that. When I got to this way of writing things, the multilevel way of writing things it made me realize when I add in beta two what I'm trying to do is explain some of the reasons people are different at time zero. When I add in beta three, the group by time interaction I'm trying to explain some of the reasons people are different in terms of their slopes.

So here are the estimates. Let's take a look first at the fixed effects. Now we've got the intercept and the slope estimates. Those are for the group equals zero. That is non-endogenous group. So the non-endogenous group starts off at about 22 ½ and they decline by about 2 1/3 points per week. The intercept difference between endogenous and non-endogenous is about two points. The endogenous patients are about two points higher and that's marginally significant. The thing of most interest here is the interaction, endogenous by time interaction, beta three. That couldn't be even more non-significant. It's like zero and highly non-significant, so what that is saying is that these two groups they differ by about two points and that difference is about the same at every single time point. There is absolutely no evidence of group by interaction, no evidence that the rates of change across time are any different between these two groups of patients.

We get that from the fixed effects. Now let's look at the random effect variances and what we see there. Now remember I said if beta two was significant or important let's say then the intercept variance would be reduced because we'll have explained some of the reasons people are different at time zero. Well look what happened. The intercept variance was at 12.63. Now it's at about a point less. Why? Because we've explained some of the reasons that people vary at time zero by knowing what group they're in. Now admittedly this is not quite at .05, but you see the direction of things. We have done some explanation of why people are different at time zero and therefore there is less individual variation. We've explained some of it. What about the slope? The slope variance was 2.08. What is it now? 2.08, why? We've explained nothing. The group by time interaction, knowing what group they're in has explained nothing about what is their slope, so you see that the group by time interaction is highly non-significant. You also see that the slope variance hasn't changed one bit. Those two things are congruent. They're telling you the same thing. Basically there is individual variation in the slopes and knowing what group they're in tells us nothing about that variation.

Does this model fit better than that last model? Well- Yes.

Question: Yes, in this case the interaction does not have any significance, so is that better go back to the model without interactions ****? Is that better or just to use basic model?

DON HEDEKER:You're exactly right. The question is in this model we've now seen that there is no group by time interaction. Should we stick with this model or should we simplify the model? Yes, I would advise simplifying the model at this point and taking out the group by time interaction and if you do that actually what you'll find is that the endogenous intercept will now represent what is the difference between the groups at every time point because without the group by time interaction in the model the group effect represents the group difference at all time

points and it will again be estimated to be about two, but this time it will be significant because it's not just at time zero, but at all the time points. So you're exactly right. Typically if you have interactions in a model that are non-significant you simplify the model in throwing out those non-significant interactions. That's right.

But let's compare this model to the previous model that didn't have beta two and beta three. We can do a likelihood ratio test comparing the two and we find that that's non-significant, so knowing what group they're in is not really an important thing per se, which is kind of interesting. Again what this suggests is that the drug does not work necessarily differentially for these two groups of patients. There is no evidence that it works differently for the two groups.

And here is the fit of the individual trend lines for endogenous, non-endogenous and the observed means across time. Again, there is about a two point difference at every single time point, so that is observed at baseline as well as all follow-up time points, so there is no evidence that they have different rates of change, that the drug works differently for the two groups. This is then how to bring that into the analysis. Again, it's really easy in PROC Mixed. We just add in the endogenous, the grouping variable, the endogenous by time interaction and so here we have a random trend model with the group effects.

Well why don't we try to run and estimate some of these models before we get into the next bit of material? So let's go into SAS and run these examples that we've considered this morning? So we'll go to start and hopefully find SAS. Here it is, wonderful, 9.2. The power to know. Yes, question.

Question: **** showing the slid you put end week. Did you add that variable?

DON HEDEKER:That variable was in the original dataset. It was the product of group by time, but you could actually in SAS you could have that as group times time. You can do it that way too. It's the same thing, but it actually was a variable in my dataset.

Question: We have **** from this ****. That one does not ****.

DON HEDEKER:I thought I did it from the start menu actually.

Question: You have to go to start and then the regular menu, not the one that tells you ****.

DON HEDEKER:I went here.

Question: Choose the English, not the ****.

DON HEDEKER:Okay, I'm sorry, okay there and then additionally here.

Question: The bottom.

DON HEDEKER:The bottom, okay, I'll try that.

Question: Now we have to choose ****.

DON HEDEKER:Yeah, I see. Okay, great, thanks, fantastic. The power not to know. So I imagine you discovered this yesterday.

Question: To me it always works from run, SAS. That should work.

DON HEDEKER:That should work, okay.

Question: One, two, three, go.

DON HEDEKER:Right, go some more.

Question: You can do it.

DON HEDEKER:Right. You can run it.

Question:Yeah, but the **** SAS. We just did ****.

Speaker:On the file.

Question: The file itself, yeah. It's already restored.

DON HEDEKER:Okay, let me try that.

Question:Just give it a second. Which one were you ****.

Speaker:Risbee [ph] M, I think is the one that I want. Yeah, that is what I want.

Question: Let's try that. There you are.

DON HEDEKER:It works much better on my abacus at home.

Question: Is that a multilevel abacus?

DON HEDEKER:Right, I'm very fast with it.

Question: It opens the **** now. I'm going to try it one more time ****. It's actually opening the guide. I would wait for this one because it opened ****. I'll have to wait for that one. If not, I can switch computers. I have ****. Yeah, that is what I did. He has it set to open the guide. Mine does the same thing. Okay, one more time and I'll bring in my laptop.

Question: The one thing we haven't done is close everything and do it the start English way. The last time we did start English way the other one was already locked up and open. You got the **** there, so we have not replicated the one way that worked last time.

DON HEDEKER:That one.

Question: Yes.

DON HEDEKER:Thank you. It seems like that should have worked, right. We have a pain in our SAS. It can't take this long, right. So presumably that, right. That is what we just tried I think.

Question: Is there any other language ****?

DON HEDEKER:Do I know any other languages? I was once I had to do a training in Holland, in the Netherlands and the program was in English, but the operating system was in Dutch. It made for quite an interesting thing. It's like luckily I know a little German and it's kind of similar, but they spell things a bit different.

Question: Get Eric just in case I can't pull this off. Tell me if it shows up.

DON HEDEKER:It's searching.

Question: Can you do the monitor thingy?

Question: What is that?

Question: Can you do the monitor thingamajig?

[Crosstalk]

DON HEDEKER:I'll need that one again when I go back to presentations. Thank you. Thanks very much.

Question: And then which one? This is the one you were using, right. I got it on the ****.

DON HEDEKER:Yeah, the problem is I have comments in my-

Question: Sorry, here.

DON HEDEKER:Yeah, it's okay. Okay, thanks.

Question: Is it on that flash drive?

DON HEDEKER:I have what I need on another flash drive. It's fine. That's fine.

Question: We'll get this one unfrozen and then she can maybe swap you back at a break or something.

DON HEDEKER:That sounds good, thanks. All right, I think we're good to go. Thank you very much. I appreciate it. Usually I have to mess with that and it takes a lot longer time and it's not as interesting because if you're doing a presentation and you're software doesn't work it's really not a good place to be.

So now we're in SAS presumably in the English language version 9.2, something like that, and what we want to run are these Risbee datasets that we've been exploring this morning. So again, this syntax is exactly what we've seen in the programs, the PDF files. Here first is the random intercepts model that I didn't actually run, but I used as a test and here note that we only have random intercept. Here is a random intercept in time model, then comes the random intercept in time model with endogenous and endogenous by week. Then I also have a model considering curval [ph] linear time effects, which I haven't yet spoken of and that's basically it.

So many of these runs are things we've done before. What I need to do now though is to change the location of that data file because as I believe where it is, is where- It's on C training.

Question: Or also just **** C training ****.

Question: What happens right there. I mean well go one up.

DON HEDEKER:One up, okay. C training and then this name here. Can I copy the name, do a control C?

[Crosstalk]

Question: Double **** that and then it will appear on your toolbar.

DON HEDEKER:Sure, sure, right and then I can- Can I copy that doing a control C? Okay, that will make it a little bit easier. Thank you. My abacus doesn't require me quite so much. Hey, that worked, all right, okay, great. Thank you.

So all I've done is to change location of that data file and then I'm going to be running these various runs, so if you want to do it with me we can all do it at the same time or if you want to do it later. We can have a race; see who runs his model quicker than I. Of course at this point all I've got to do is hit the running person icon and it will run all of these commands, so let's go through that. You already ran that. Okay, let's see what happens. So it's running, PROC Mixed. So it looks like it went through. It got through four mixed runs, so let's go to the top and interpret the output that we're getting from PROC Mixed.

The first run I've got is a random intercept model. The reason that you can see that is in the dimension statements it says columns in Z per subject. Remember Z is a random effect design matrix. It says that there is one column, there is one random effect. There is 66 subjects. There is 66 observations per subject. Covariance parameters is just two because there is just between subject variance, within subject variance in this simple model. Columns in X two. There is only an intercept and time. That is it in this first run. There is 66 IDs and it lists all of these out. If you don't want to see these IDs remember yesterday I think on one of examples I had no CL

print or something like that as an option, then it doesn't print these different values of the IDs out.

So let's go down and see some additional information. Number of observations, this particular dataset and maybe we can take a look at it just so I can give you a sense of what it looks like, so let's open with select program from a list. I like using Notepad. That is always my favorite. Here is the dataset that we're actually reading in. It has as its first field the ID. The next field is the Hamilton Depression level. The third is this values of ones that we don't need for SAS. The next is the value of week zero, one, two, three, four, five. The next is endogenous, non-endogenous and then the final one is the product of group by time, endogenous by week. Now you'll notice that the way this dataset has been created every person has six lines and if they're missing at a given observation like 106 is here for example they have the SAS missing value code, so that is why in the output we're going to see is that it reads in 396 lines, 66 times 6, but it doesn't use all of those in the analysis. It throws out these observations that are missing. Now I want to make clear it doesn't throw out the subject. It's throws out the particular observation, so for example, subject number 106 is going- These observations will be in the analysis. This one won't because we have no information on the dependent variable of that time point.

107, that person was missing at week- at this particular week, week four. Number 113, again was missing at the last time point. 114 alternatively is missing at the first time point, so it's going to use whatever data it has. It won't throw out the subject, but it will throw out the missing observations using the available data it has and in total out of those 396 records there is 21 instances of missing observations, not missing subjects, but missing observations. So that is why when you get to that part of the output what you see is what you get. 396 number of observations used, 375, number of observations not used, 21. That is not 21 subjects. That is 21 observations.

Then this very simple model didn't take very many iterations to converge to a solution. Here we get the results of the analysis. This is a random intercept in time model. What they label as one, one is the random subject intercept, the variance of that, so that is the between subject variance, this estimate of 16. The residual is the level one or error variance, residual variance 19. Now let's jot down this likelihood value, so we can replicate the likelihood ratio of tests that we did on the slide earlier and this value of the deviants here, minus two times the log likelihood is 2285.2, so I'm going to jot that down, 2285.2. So we'll compare this to the subsequent model. Now in this model in terms of fixed effects we have the intercept estimate of 23 ½ and the slope estimate of about negative 2 1/3.

Our next analysis now has same number of subjects, 66, but now has 4 covariance parameters and 2 columns in the Z matrix. Why four? Random intercept, random slope, intercept slope, covariance, error variance, so those are the four. We get down here to the deviance, 2219.0. Is that right? Did I read that right? So doing that bit of subtraction you get a value of 66.2. That is the comparison between these two models and a test of the null hypothesis that the slope variance and slope intercept covariance, that those two parameters equal zero because the first model didn't have those two parameters. This model does. So this is a highly significant result that is again saying that people vary in terms of their slopes.

Here is the estimate of the slope variance. Here is the estimate of the slope intercept covariance. Again, the way it labels it is unstructured matrix first row, first column that corresponds to the intercept variance, unstructured matrix second row, first column that corresponds to the intercept slope covariance, unstructured matrix second row, second column. That corresponds to the slope variance, the time variance. You can see that when you ask **** for the estimated G matrix it prints it out in that particular manner.

Likewise, the G correlation matrix shows the covariance in correlational form. That is what you get with the G core option, so you can look at this covariance and when you see the covariance it's hard to know what associations does that represent. Well in terms of the correlation it's minus .28.

Now here you get these wall [ph] tests of the variance, covariance parameters because I asked it for cove [ph] test in terms of an option. Now I've said yesterday is you don't want to rely upon these wall tests here for the variance, covariance parameters. Better to use the likelihood ratio test. So you might ask why did I ask for cove test as an option. Because by default if you don't put cove test what SAS will do is it will list these estimates, but it won't list the standard errors. It won't list the Z values. It won't list the P values. That is what it does by default, but I've asked for cove test even though I know these tests are not really statistically good ones.

Why? I like to be reckless and dangerous. No, when your model is not well estimated where do you often see that showing up? You see it often with the standard errors of parameters. Standard errors are basically how certain or uncertain do you have a fix on that parameter and if your model isn't well estimated those things can either show up as being like very close to zero or incredibly large, so when you see standard errors that are very unusual that indicates that you don't have enough information to estimate the model you're throwing at it and by default if you don't put cove test you won't see these standard errors. It will just show the estimates.

I think looking at the standard errors is an important aspect in model diagnostics to see whether or not your model has been properly estimated and so I always use cove test, not so much to get these P values because I know that they're not really good tests, but to take a look at the standard errors to see for example are any of them zero or very close to zero. If you have problems in estimation you'll see something like that sometimes is that the standard error estimate will be like zero or very close, something like E to the negative 11th or something. If you see that you don't have enough information to estimate that parameter. You need to take it out or if you see a standard error that is like 9,999, some incredibly large value that is also not good information.

So here I just take a look at those just to sort of see in relation to our estimates do they look somewhat plausible. I'm looking for gross things, nothing subtle here. Is it close to zero? Is it incredibly large? None of these are problematic, but that is why I put that cove test option on is just to see whether or not the standard errors look somewhat plausible. Again, I'm not looking for anything really that subtle, but some really gross things that will happen in some cases when you don't have enough data to estimate your given model.

So that is why I put that there and then below this are the intercept at week again, just like what we saw before, but now admittedly the standard errors are going to be a little bit different

because I have a different variance, covariance structure, one that has random intercept in time effect.

So that is the model without group. Then the model with group in it and group by time interaction is the third model that I estimated. Now here when we look at this and the results again in terms of the variance, what we see is that intercepts variance has gone down a bit. It was like 12.6. Now it is 11.6. The slope variance is virtually the same. The deviance has changed from 2219 to 2214.9, so if we do a likelihood ratio test that is a value of 4.1, two degrees of freedom because I've added in endogenous and endogenous by time interaction that is not really significant, so that is not really an important improvement to our model and we see that when we look at the fixed effects and their estimates. Here we get the estimates, standard errors, T value for the fixed effects. Again, the endogenous difference at time zero is pretty small. Sorry, it's about two points and marginally significant. The endogenous by time interaction very small and highly non-significant, so there is no evidence of group by time interaction as was mentioned. What we might want to do at this point is go back and run the model including the group effect, the endogenous effect, but not the group by time interaction, so in fact, let's- maybe we should that just for the sake of example.

So I'll go to that, so that is my third mixed run and what I'll do simply is to take out the product of endogenous by week, so now I've got like a main effects model and a model with time, week, group, endogenous. What I'm saying here is that there is a consistent group difference at all time points because what I've learned is that the group difference doesn't vary with time by the last run, so now I'm going to simplify it and say well there might be a group effect that is the same across all time. Let's take a look at that.

Now to run this so I can just take this, highlight that bit of syntax and then hit the running person icon.

Question:**** need run.

DON HEDEKER:I need run, yeah, thank you. If you want to run you have to say run. Let's try that. Here is the result of that. Now first of all, if you look up at that the deviance of this model 2214.9. It's virtually the same as the last one. Why? That group by time interaction was doing nothing, virtually nothing and here are the results of the fixed effects. Now notice this. In that last model with the interaction present the endogenous main effect was the endogenous difference when time equals zero. Now I'm getting the endogenous difference at all time points. Again, it's estimated to be about two, but here is it significant, so in some ways this is a better model. It's allowing for a group difference that is consistent across time because we examined and found that it was not varying with time and now it is below the .05 level. Don't you often get-? I always get when people come in my office they have a result that is not significant. They want to make me find it to be significant. I always tell them if you want .5 to go to .05 in math things got to equal out, then you got to add a zero on the left-hand side of the decimal place to my hourly pay. It's got to equal out. So if you take away this it's got to- No, just kidding or am I?

So here there is a consistent endogenous effect. Now the last run that we didn't have a slide on, but we have the results on is the quadratic trend model. Let me say a little bit about that before we get back to the material. Endogenous, week, week squared, I'm sorry, not endogenous, Hamilton equal week, week squared. Now remember my comment when I looked at the means across time and they fell on that line and I said it looks like Mendel [ph]. It's like it's a perfect linear fit. It's like textbook data, so why would I consider a quadratic trend model? Well I did it in a way to show the students when I first was doing it how to run a quadratic trend model and I thought well there is going to be nothing here, but I'll run it just to show you here is how to run a quadratic trend model. We add week squared here. I've also added you'll note, on the random statement. That is how one can run a quadratic trend model and notice that in this syntax I'm creating interactions kind of on the fly in SAS. You can do that, week, week times week, week times week.

Now and this doesn't have group in it, but it has week and week squared. Now what we had as our deviants for our model, let's see, model two was 2219 and here the difference between model two and four is the addition of quadratic time trend on the model statement, quadratic time trend on the random statement. I've seen the means across time. I have no- I don't think that there is going to be much improvement a model fit here compared to here because it looks like the change across time is linear.

Let's look at the output there. So that is above this one that I just created and ran. I hope I'm going the right place. Am I going too far? I think so. It seems like it. No, here it is. Here is it. Yeah, here it is. 2207.6, so I jot that down. Let's say 2219 and that was the random trend model. 2207.6, I do my little bit of subtraction. I get .4. I get 11.4, chi-square 11.4. How many degrees of freedom? There is three additional variance parameters here and one additional fixed effect with four degrees of freedom. Guess what? It's significant. If you look at a table chi-square with four degrees of freedom I know all the **** tables like I do, it's significant. Now that is a little bit of a paradox, especially if you look at here is the variance, covariance parameters. Here is the fixed effects estimates down here. Week times week pre-zero, highly not as significant, so beta two is clearly not significant, but this model fits better than the previous model.

Why? Imagine this. I don't have a board, but I'll draw a picture. Let's say we have this as our average trend across time, just a straight line. I should have drawn it down because that is what we have currently, but it's the same idea, a straight line. Now imagine we just have two subjects that look like this. The trend on average is linear. The trend at the individual level is curval linear. It's quadratic. What I've done is to draw kind of like- Sorry I didn't make that- Actually let me go back and make that- So this is S1. This is S2. It looks like a lip or something like that I guess. Subject one, subject two, at the individual level you could have quadratic trend, at the average level linear trend. In a way that is what I have a bit of here because at the average level there is no affective curval linearity, no quadratic trend, but at the individual level the model actually fit better and in fact, if you look at the variance parameters there you'll see that the quadratic variance by this admittedly dubious wall test, but nonetheless is .02. What are the covariances? Close to .05.

Now does a quadratic trend make sense for depression values across time? Generally yes. It's not such a- There is some people that are going to improve and then kind of level off or there might be some people that it takes awhile for them to improve and then they go down. Actually to some extent I think truly with data like this a cubic trend might make the most sense because often it takes awhile for the drug to work and then they start to improve and then because there is a floor of rating it kind of levels off, so if you had enough time points perhaps a cubic trend would actually make sense, but clearly we're seeing some evidence of linear- of quadratic curval linearity at the individual level, not in terms of the means. The means across time largely linear, but in terms of at the individual level there is some evidence of curval linearity.

Now I don't want to oversell this. When we went from the random intercept to random trend model the chi-square difference was like 66, so adding that in was a pretty big step. What we've done now is go from random trend to random quadratic trend. The chi-square difference was like 11, so it's not like everybody is like this, but there is some evidence and it's fitting better than the linear trend model.

Now when I first ran this I couldn't believe it. I was like, what, how can this be? Like why? The means are linear, but my model is not just about the mean change across time. It's about the individual change across time and then I thought of something like this. It's like yeah, just because the average trend is linear and maybe I should- doesn't mean that the individual trend is linear and actually if you go back to that first plot I had, the spaghetti plot you'll see that there are some people that sort of illustrate a curval linear kind of pattern across time, so adding in the quadratic here in terms of the random effects did improve the models to some degree.

Any questions about this syntax or the output you get from SAS running PROC Mixed? Yes.

Question:**** it seems like a lot of behavior or biological as shaped in humans should be **** time observations. Is there a threshold in terms of like **** study design or ****?

DON HEDEKER:Right, so the question is with behavioral outcomes an S-shape kind of curve would make sense, but you need a certain number of time points to estimate an S-shape curve. What sort of should you think about? Well let's think real basic. If you have two time points all you can get is linear change, so if you suspect that you're outcome changes in a quadratic or curval linear manner, something like that a study with two time points will not be able to detect that. Three time points the most you can estimate is a quadratic, so if you suspect S-shaped three time points is not going to get it. There is no way. It's impossible. So you'll need at least four time points if you suspect like an S-shape kind of curve to your dependent variable across time, four or more basically.

Now in this study I do have- I have up to six, so looking at the cubic is actually possible. We could actually run that. Yes.

Question:**** have a difficulty to interpret the degree of freedom in the fixed **** solution, fixed effect ****?

DON HEDEKER:For this particular model?

Question:Yeah, ****.

DON HEDEKER:It's actually down isn't it? Isn't it this stuff?

Question: Yes.

DON HEDEKER:Right, degrees of freedom.

Question: No, base one, other one.

DON HEDEKER:Probably with endogenous and-

Question:**** just like 2.3 ****.

DON HEDEKER:Okay, 243.

Question:243, for the ****, yeah.

DON HEDEKER:This one. Right, so first of all, there is 66 subjects, but there is 275 number of observations in total because there is basically 66 subjects times 6 is 296, but there is 21 missing observations, so the affective number of observations is 275. So again, what SAS does and I don't perfectly understand it is it uses the number of observations and also how correlated are the observations and then it also distinguishes between whether the variables that time varying or time invariant, covariate to determine the degrees of freedom. Now you might ask why is it.

Question: My question is the difference between the weeks and the ****.

DON HEDEKER:I don't know to be honest.

Question:Yeah, for the weeks it seems like it treated weeks as a subject.

DON HEDEKER:Yeah, you're right. You're right. It does seem kind of illogical. Week is actually a variable that changes, but endogenous does not and yet week has less degrees of freedom than endogenous. I don't know, but let me again state the difference between this and this is really it doesn't matter because if you look at the P values under 65- Look at the critical value under 65 or 243. It's not going to be very different. So again, I don't have a good answer for your question. How does SAS determine these degrees of freedom? I know the elements that go in it and I know something about it, but I don't really know why it's coming out like this, but the ultimate answer is that it doesn't really matter in this case because there is plenty of observations and the difference in P value between using 65 or 243 is miniscule. It's not an important difference. Again, if you are worried about this the most conservative way of calculating degrees of freedom the **** Roger approach, that is an option that you can select as one of your options in SAS. Any other questions about this?

Well before we break for lunch I want to get into the drug levels of this study because this study was aimed at looking at whether drug influences the outcome. It wasn't really aimed at looking at endogenous with the outcome. That was just part of it. What they're really focusing on is looking at the drug level, the measures of Desipramine and Imipramine and their relationship to depression illness. So let's go back to the thing here. So time variant covariates within subject and between subject effects. This is covered in a section of our book, section 5.2 in fact.

So I'm going to ignore endogenous and non-endogenous. That really wasn't an important aspect. I could put it in here as a main effect. We saw that that was at .04 per se, so that might be useful, but here I'm going to focus on the effects of drug. Now again I'm writing it here first in this matrix kind of way. Now here one thing to realize is that in this particular study there we six time points, but the first two time points there was no measurement of drug. That was a drug washout period, so once I want to treat drug as a covariate I'm really only focusing on the last four time points because at that first two time points there is no information about drug because that is when there was drug washout. So in this part I'm only looking at the last four time points in the study and so therefore, time here will go from zero, one, two, three.

Now you'll notice I have IMI and DMI in the model and I've got logs for IMI and DMI. Why do I have logs for DMI and IMI, anybody have an idea? If you don't, don't worry. It took me a long time. I'm transforming IMI and DMI by logs. Why am I doing that? Here is why? Let's imagine that we're looking at the relationship of income to happiness. So I get all of your incomes. I get all of your levels of happiness. Happiness is my dependent variable, so I get something like this. Let's see, happiness. Does money buy you happiness? Age old question, right. So say I got a bunch of data points like this. Happiness is the dependent variable. Money is the independent variable. Now suspend your belief for one second. Imagine through this door either Steve Jobs or Bill Gates walks in or maybe they both walk in together holding hands. That would never happen, right, but just imagine that happens. Let's say one of them. Now where are they going to be in this plot? They'll be in the next room, right, in terms of income. They're way out, way out. So where they are, say they're over here. Let's say they're having a particularly bad day. Stocks of Microsoft are down. Bill Gates comes in. He has got a lot of money, but he is not very happy. What is going to happen to our estimate of the slope in this context? It's going to go like that. What if he had a great day, he was up here? Same level of income, but he is much happier this day. It's going to go up there. In other words, observations that are out there in terms of X have potential for changing the slope, so it's called influence, so any observation that is far out in terms of X has a potential to change that slope to a much greater degree than observations over here.

When I looked at the distribution of IMI and DMI I saw some observations that were kind of out there, high values and I wanted to bring in that potential influence of those outlying values. That is why I took a log. Log transformation is great transformation for bringing in high values, so I didn't do it because there is some assumption of normality for IMI or DMI because there is no assumption, distributional assumptions for the regressors.

In this kind of model we're not assuming any kind of normal distribution for X variables. We're assuming normal for Y and most precisely it's conditional distribution of Y given X that we're assuming normality for, but there is no assumption regarding distribution of X. However, if you

are having as a regressor at X variable and there is one or there is a few observations that are kind of out there, high values they have the potential for changing the slopes to a greater degree than these observations down here. You want to potentially bring them in. That is why I took a log, so I took a log of IMI and DMI for that reason, not for any kind of distributional reason and to be honest with biological measurements like this it's pretty standard to use a log transformation rather than the milligrams per deciliter or however they were measured.

So I ran this model and when I first rant it I wrote it this way and proceeded, but here let me show you the multilevel way of writing this model. So here I've got the multilevel way of writing the model and now I've got three level one predictors because time changes value across time. IMI changes value across time. DMI changes values across time. They both, all three have subscripts I and J. They change values across time and now all the coefficients come down at level two to be modeled in terms of subject level variables. Now in this particular run I don't have any. I could have included endogenous perhaps as a subject level one, but I- considering a random subject intercept and time trend and then I see this. When I write it this way, in this matrix way I don't quite realize the same thing as when I write it this way. I'm not allowing these to be random. What does that mean? I'm saying that there is an overall effect of the drug level and that it doesn't vary from subject to subject. There is just one effect of IMI. There is just one effect of DMI.

Now could I allow these to be random? Sure. What would it mean? It would mean that the relationship of drug to the outcome, the depression level varies from subject to subject. Does that make sense? Yeah, it probably does actually because they know in psychiatry that these drugs don't work for every person, so there is already some knowledge that the drug response relationship varies from subject to subject. Yes.

Question: Wouldn't that better account for those outliers that you were talking about earlier?

DON HEDEKER:Yeah, it could. It could too, yeah.

Question: Where the person didn't have any response.

DON HEDEKER:It could, yeah. You're right. It could account for that, so the idea of including these as random effects makes pretty good scientific sense actually. Now here I haven't done that. One thing to consider is this. I've got 66 subjects. I've got four time points basically in this part of the study and considering a model with four random effects it's a little bit much with this amount of data.

Let me give you an example of where I did do this exact kind of model where I had more data and was able to run this kind of thing more successfully. Are some of you familiar with the Fishbein and Ajzik theory of reasoned action? No, okay, a little bit. It's just an idea in behavioral sciences that Fishbein and Ajzik postulated I guess in which they said that behavior is caused by behavioral intentions and behavioral intentions is influenced by a person's attitude towards the behavior and their subjective norms about the behavior, so what they said basically is behavioral intentions influenced by subjective norms, kind of what you think other people think about that behavior and your attitudes towards behavioral, what you think about it.

Well I had an adolescent dataset in which I had like 1,000 subjects measured over several points in time and then there attitudes towards behavior and subjective norms were time variant covariates much like IMI and DMI and my outcome was behavioral intention and this was smoking, behavioral intentions of smoke, adolescent subsample and I looked at whether or not the relationship of subjective norms with intentions, attitudes towards behavior with intentions, whether that varied from subject to subject and of course it did. There was significant variation there in terms of these time varying covariates because basically in an adolescent sample there is very great individual heterogeneity in the relationship between attitudes and intentions, subjective norms and intentions, so that kind of model works kind of well in some ways, but you need a fair amount of data to estimate it.

Now in discussing things with you it seems like a lot of you have pretty big datasets, so that possibility of treating time variant covariates as random at the subject level is certainly something that you could consider and again, what it would get at is whether or not there is heterogeneity across subjects in these relationships. So here it makes logical sense. There is not a ton of data by which to estimate for random effects in a dataset that has only got 66 subjects, but in datasets that are larger it certainly can be done.

So I ran this model and here are the interpretations of it, the parameters. Beta not is average weight to Hamilton for drug-free patients. Drug-free, what I mean by that more precisely is for patients that have levels of log IMI and log DMI equal to zero, so drug-free is kind of an incorrect phrase there. Beta not of course is always the intercept when all X equals zero. It's the value of Y when all X equals zero. Beta one is the effect of time. Beta two is the average Hamilton per unit change in log IMI. Beta three is average Hamilton difference for unit change in log DMI and then I've got a random intercept in time and here again at week two is the actual study week one week after the drug washout period, which is coded at zero in the analysis at the last four time study points. So I'm looking at the last four study time points. I'm coding them zero, one, two, three, so what is called week zero in this analysis is physically week two in the study.

Now here is what I got when I ran this random intercept in time model with the time variant covariates. So I look at the results. The intercept estimated to be 21. The slope is minus two, insignificant. People are improving by about two points per week. IMI is positive and non-significant. DMI is negative and only marginally significant. So first of all, what is the direction that you would suspect of the drug effects? What would the psychiatrist want to see there? Negative, negative, as drug goes up depression goes down. Well I'm seeing that on DMI, the metabolite, but it's only .06. IMI, it's positive and .48. When I saw that result I was scratching my head and you can see I scratch my head a lot because I lost a lot of hair in doing so. I couldn't understand this.

Let me give you a little bit of background on this. When Risbee published this study in '77 admittedly that is a long time ago with regards to longitudinal data analysis methods, what him and his colleagues did was this. They took the Hamilton at the final time point, at week six and they trichotomized it in terms of was the person what they called not depressed, mild depressed or still depressed, so they used cut points, arbitrary cut points of the Hamilton score at week six.

They classified people in terms of this trichotomy. They then took the drug plasma level at the last time point, IMI and DMI. They added it together. They then did a Spearman correlation between the sum of the drug levels and the trichotomized Hamilton at week six and they got a significant result. They did something very crude, very basic. They got a significant result.

Now you've been here for two days learning mixed models and all you get is .06. When I ran this I was perplexed. Why am I only getting .06 when what they did was much simpler, much less elegant? They got significant results. Anybody have an idea of what I'm missing here?

Question: Something to do with the logs.

DON HEDEKER:What? Has to do with the logs, yeah, that is an interesting idea actually because I don't think they took logs.

Question:**** considered the time variable for the drug ****.

DON HEDEKER:Right, I'm treating it as time varying. They only looked at the last time point. I mean that might play a role here. Maybe there is more of an effect at week- at the last time point and I'm looking across all four time points. Yes.

Question:**** the drug has to biologically build up in the body, the metabolites and so there would more of an extreme at the very end.

DON HEDEKER:That is true. That is true. I mean maybe that is one aspect is that I'm looking at these four time points and maybe the effect isn't really there in the beginning, but it develops across time and in this model I'm including kind of like no result with some result and getting kind of like in the middle. That is true. That is possible. Another thing to consider- The first thing that came to my mind was I thought well I've got these two predictor variables. They're highly correlated with each other and they're wiping out each other's effects. That is the first thing I thought of at least. That wasn't it though actually.

Question: Are they still held constant and ****?

DON HEDEKER:Yes, yes, this is just like a regression model, so you're estimating the effect of DMI holding constant the effect of IMI and right, if these two variables are highly correlated they overlap a lot then they're independent effects are virtually nil because you can't estimate independent effects of two variables that are highly dependent. It's just logically you can't do it. I thought of that first. It wasn't the case necessarily that they were that correlated with each other.

Well what I went through was you're right that in some ways looking at across all time points maybe that is not the best way to go, but I was thinking about is this. What exactly is beta two? What exactly is beta three? What am I doing here in this model? What does that represent? And I realized looking at this what I'm trying to get at is at a particular time point is there an association between drug level and the outcome. I mean what beta two really represents is like you regressed the Hamilton on IMI and DMI and you did it four times at every time point and

then you averaged across those time points. It's getting at is there a concurrent relationship between drug and outcome averaged across time. It's sort of what is sometimes called an average cross sectional effect treated in this way.

Again, beta two is just like if I did a stratified analysis at the four time points and at each time point I regressed depression on these two covariates IMI and DMI. I did that four times. Beta two is like the averaging of the IMI coefficients. Beta three is like the averaging of the four DMI coefficients. It's an average cross sectional effect. It got me to thinking. It's was like well maybe that is not the way drug and outcome really act. Maybe it's not like at a given time point how your drug is- relates to your response. Maybe it's like how you are at a time point changes the next response or maybe how your drug is relative at a time point will influence how your depression is relative to where it was, the change in your depression, something like that.

I want to point out that when you're treating with time variant covariates there is a lot of subtleties there and just throwing them in the model the way I did like this, the naïve way, what you're getting at is basically an average cross section effect across all time points, so you're assuming that basically at a given time point there is a relationship between these variables and then just saying, averaging that across all time points. Maybe that is not the way it operates. As has been suggested another thing that could be looked at is interaction with time, so maybe that relationship starts off as null and then grows. So there is lots of possibilities.

What I eventually did and what I'll show you next is this idea of maybe the drug level is not related to the depression level at that time point, but more precisely where it's at, at that time point relative to where it was, in other words, the change in the severity. So I looked at that in two different ways here and here are the results of that analysis.

In the first analysis on top what I did was to included as a covariate their baseline level of depression, so now my outcome variable is adjusting for where they were, where they started and I'm looking at the drug effect on their severity score at a time point, but adjusting for how they started out. On the bottom what I did was express each of the four time points, depression levels as change scores relative to baseline, so I'm modeling on the bottom change from baseline in the depression score and is that related to drug level.

Now when I did this either way I was able to save my career because I got significant results on DMI and IMI, highly significant results. Still not significant results on IMI Imipramine and this is kind of an interesting result. The parent drug is Imipramine, but in your body it metabolizes to Desipramine. It's the metabolite that is highly significant. It kind of suggests in some ways the metabolism is perhaps what is related to depression because it's the metabolite, the level of metabolite that is really driving this relationship.

Now in their analysis, in Risbee again they summed these two things together. I always thought that that was kind of a weird idea of summing the parent and the metabolite together and this kind of bears out that the summation glosses over the fact that it's really the metabolite that is driving this relationship, but what the metabolite is really related to is not so much the level of depression at the time point, but the level of depression at a time point relative to where they were. So I'm looking more at change in the Hamilton and is that related to the level of the drug.

Now I've presented this several times and a lot of people suggest well maybe you should look at drug level and Hamilton at a later time point, kind of a lagged model and I think that makes potentially some sense too that maybe it's the covariate level of drug is related, not to the Hamilton at this time point, but next time point, something like that and that again is something that one could investigate as a possibility.

Now when I did this I was trying to think of how can I create descriptive statistics that- simple descriptive statistics that helps me explain these results because if you want to present to a group of psychiatrists- Any psychiatrists in the audience by the way? Okay, I can badmouth psychiatrists. No, I'm kidding. But if you're- to more lay or **** audience you want to present these kinds of analysis in conjunction with some very simple statistics to get a feel for what the results represent.

Well here it's a bit of a challenge because with endogenous let's say, that was a grouping variable, so I can easily plot the means across time and that can easily show how the groups are different in terms of their trends across time or not, but here my covariates are continuous variables and they're changing across time. So what kind of descriptive statistics can I use to help me explain these estimates from this mixed model? So think about this. You've got a dependent variable that is changing across time. You've got independent variables, two of them that are also changing across time. What kind of simple statistics might you do the help explain this? Let's say forget about time for the moment. Let's say you only have one time point.

Question: Correlations.

DON HEDEKER:Correlations, yeah, that is what I thought. I'll just do some correlations. Do correlations of the drug level with the Hamilton at the different time points. People can understand correlations and that is kind of a useful way to express these results. When I did that it helped me also understand why my first attempt at this modeling was unsuccessful and this attempt was more successful because here is what I did. I took the Hamilton total score and first I correlated it with IMI and DMI at the different time points. Now this bears out exactly the comment that we saw earlier, heard earlier that well maybe it takes awhile for the drug affect to develop. If you look at DMI here week two there is a negative correlation. It's really not significant. Week three negative, not significant, then it develops and in terms of IMI it's negative, but largely non-significant across the board.

Now this set of results was akin to my first attempt when my first attempt was just treating this as time variance covariates and again I said basically beta, those betas were like average cross sectional results, so my DMI effect that was significant, .06 was like an averaging of this, this, this and this. Two are significant. Two are not. That is why I got this kind of middling result, not very powerful.

Then on the bottom I have the same correlations, but with the change from baseline, what I subsequently went to. Now there, interestingly, even DMI even at week two this is kind of interesting because it is true that there is this notion that it takes awhile for the drugs to work, but even here at week two that is like they did it at the end of the week, so they've only been on

drugs for about six days. At the end of the week even. There is a strong relationship with the DMI level at that point and their Hamilton change from baseline, not so much with their Hamilton level at that point, but their change, their improvement at that time point and that is rather consistent for DMI across all time points, a very strong and negative effect, higher level of drug, greater change score, greater negative change score.

Now IMI again comes out hardly at all, so if IMI is really not driving things and again when they did their analysis they combined these two and so I'm they were fortunate in that they looked at week five. If week five things are starting to happen either way presumably, but what they didn't uncover is that it's not the parent compound, but the metabolite that is really driving the relationship. Yes, question.

Question:**** why it goes from 3.5 **** down during week 3 and 4.

DON HEDEKER:Yeah, I wouldn't make too much of this. It's not really that big of a change per se, but yeah, at first it's like eh, but yeah. This is not really that big of a difference per se. I would tend to view these all as being pretty close to each other. Yeah.

Question: It seems IMI and DMI are so much correlated.

DON HEDEKER:They're not that much correlated with each other. They're not. They're not. They are not. That was my first thought. I looked at that. They're not that highly correlated actually. In some ways you would expect them to potentially be negatively correlated because if there is more IMI that would mean there is less DMI floating around. They are to some degree, but it's not like- it's nowhere near like 1 or .5 or something like that. They're correlated, but not to a tremendous amount.

Question: Because I was thinking to separate it and ****.

DON HEDEKER:Yeah, I mean that is the first thing I looked at to be honest and that they're moderately correlated, but not near dangerously so. It was not that big of a deal.

Question: Excuse me. You're talking about IMI and DMI.

DON HEDEKER:Yes and if they're associated because again in all of these models I'm putting them both in, so when you have correlated explanatory variables and they're highly correlated they can wipe out each other's effect because in the model you're trying to estimate the independent effect of the variables. What is the effect of this in controlling for the other ones? And if you have two X variables, two explanatory variables that are highly correlated they're going to wipe out each other's effect. You're going to see virtually nothing because you're trying to say what is the effect of this controlling for this. Well you can't estimate that well because they're- you can't separate them.

Question: And for the ****, the change rate are you comparing the time to versus times one and times three versus time two or are you comparing each one for the very first baseline?

DON HEDEKER:Exactly, that is what I did. I just did change from baseline, so these four time points for each one I'm subtracting off their baseline, their first measurement of the Hamilton.

Question: Is it better to use like ****.

DON HEDEKER:The subsequent change, you could. It sort of depends. I mean you would then get at each time point what is their weekly change. Here instead what I'm saying is what is their change relative to what they came in with.

Question: Their beginning.

DON HEDEKER:Right, exactly, yeah.

Question: So if the IMI and the DMI had been really highly correlated you would have run those separately.

DON HEDEKER:Yeah, presumably, yeah and then you would have to go with one model or the other. I mean conceptually you would like to have them both, but if they're so highly correlated you don't have the data to estimate the model with both of them in really. Any other questions?

Well here is another aspect that comes in with time variant covariates that I didn't realize when I ran the results I showed you, but since then I've come the learn this aspect. This is kind of an interesting one. This is a kind of model I'm considering right now. Let's just say I have one time variant covariate just to make this idea a little bit simpler. So let's say here is have DMI for example. This is the kind of multilevel model I'm running and the question there that might come up is well is the effect of X a within subject effect. What I mean by that is, is this coefficient that I'm getting, beta two, which is the effect of X on Y here does it reflect as a person changes X how much does their Y value change because in some ways that is what I want to know here. I want to know as a person changes their drug level does their Hamilton change.

So that is what I would call a within subject effect. As the person's level of the covariate changes does their value of the dependent variable change? And I say well it's in the within subject model here XIJ, so isn't that a within subject effect? It seems like it is, but think about this. What about- So here I have XIJ, a time variant covariate. It varies with time points. Is this an equality, what I've written right here? Hopefully it is. What I've said is- Well XIJ could be written as equal to XIJ plus the subject's mean across time points minus their mean across time points, so I'm just adding and subtracting in this and then I reorganized terms and notice what I have is XIJ being equal to is the subject's mean across time and their deviation of their covariate value at a time point relative to that mean. When I put XIJ in the way I have really what I'm assuming is that the two of these have the same effect.

Now what are these two? Let's say that I'm looking at the dependent variable, level of optimism about the future of the world. That is my dependent variable. I'm modeling the level of optimism about the future of the world and let's say I get each one of your- some sort of measure of your positive effects of your happiness. In general I would think that people who are happy have more optimism about the future of the world. That would be like the between subject

effect. You measure them. They're an optimistic person. I'm sorry; they're a happy person, so therefore they're optimistic about the world. Let's say you take that person. Let's say on a given day they've had a whole lot of trouble. They're not very happy that day. What happens to their level of optimism? It might go down, so that is the meaning of this separation of the between and the within subject effect. Happy people are more likely to be optimistic across all time points. That is the between subject effect. On a given day even a happy person could be having a bad day and they may not be so optimistic or conversely a very unhappy person they won the lottery that day, all the sudden they're very optimistic. That is the within subject effect.

Now again, if you just put the time variant covariate in as a covariate the way I have I'm assuming that those two effects are the same because there is only one coefficient for XIJ here, so now I want to explore this and develop whether or not there is a differential effect of the drug between subjects and within subject. The between subject effect here would be basically if a person has consistently high levels of drug let's say across time do they have consistently lower levels of Hamilton across time and the within subject effect would be at a particular time point if a subject is higher or lower on drug at that time point are they higher or lower in terms of depression. These two effects could be different.

Let me show you a graph the represents this idea, a couple of graphs. This is a time varying covariate effect, a purely between subjects effect, so here I've got three subjects. They're each measured at two time points. Notice for every subject there is- for a given subject there is no effect of X on Y, but overall there is a positive relationship between X and Y. As X goes up Y goes up, but this is purely a between subject effect. As the average of a subject's X value is related to the average of the subject's Y value, but as the subject changes, the subject **** change X. I'm sorry, as the subject changes X there is no change in Y. Purely between subject effect.

Now here is a plot, same six data points, purely within subject effect. Here say there is two subjects. Here within a subject as they change X, Y changes rather dramatically. However, the average X and Y are not related. That is for this subject the average of X would be here. For this subject the average of X would be here. They have the same value of Y. This is purely within subject effect. As a person changes their level their outcome changes, but as you go across subjects the average of X and the average of Y are unrelated.

Now I made these plots in this way so that you could see that when you look at the data you can't tell necessarily. It's the same six data points, just organized somewhat differently. So how can we tell if it's this or this? Well it's not that hard actually. All we got to do is to create two versions of our time variant covariate, the average of X and the deviation of X at a time point relative to that average and now we can test for the differential affect of is it the average of X related to the average of Y or is it as X changes at a time point that is related to Y, the between subject effect, the within subject effect.

Now let me say a little bit about the between subject effect. The between subject effect it's just like you took- you averaged across time X, you averaged across time Y and you did a simple regression of the average of Y and the average of X. This coefficient here is exactly in meaning that relationship. Is the average of X across time related to the average of Y across time? Is it

that people that are generally happy generally optimistic? That is what that answers. The within subject effect says at a particular time point if a person is higher or lower than their average, does that influence their dependent variable value at that time point relative to its average. That is the within subject or longitudinal effect.

Now I think when we run these models with time variant covariates in terms of our mind we're probably thinking that this- as a person changes X do you see a change in Y, but when you run models as I've done already, which is just this you're really assuming the same slope for both of these ideas because I only have XIJ in here and XIJ is equal to the sum of the average plus the deviation minus the average. So the beta two is the same for these two different effects.

Well can we test this? Yes, what you do- The idea is this. We can run this model. This model assumes the between and within are the same. We can run the more complicated model that says they're different, compare them by likelihood ratio test, see which hypothesis generally holds, so let's see that with our dataset. And incidentally, this shows a little bit also of like Y, the model with only XIJ assumes that they're the same, so suppose they are the same and equal to beta star, then the effect of XIJ is beta star times the mean plus beta star times the deviation and you can see that then equals beta star times XIJ, so just putting in XIJ, the time variant covariate you're assuming this, that the two have the same effect. So do they have the same effect? We don't know a priori, but when we put in XIJ that is what we're assuming, so it could be a dubious assumption, but it's easy to test.

Now here is another thing that comes to mind. There is no guarantee that they even agree on sign because you might say well who cares if they're the same, they're probably going to be in the same direction, I'm not going to bother with this. Well let me show you this little slide that I made about that last point. In this slide I've got two observations per subject for five subjects. The within subject effect here is negative. As X goes up Y goes down. The between subject effect alternatively is positive. As the mean of X goes up the mean of Y goes up. Here I mean, mean for subject. As the subject mean on X goes up the subject mean of Y goes up. However, within a subject as X goes Y goes down. So there is no guarantee that they are even in the same direction.

In some ways you can think of this, this is sort of an example of ecological fallacy potentially. An ecological fallacy mostly you find out about it in sociology where they say they look at something at the neighborhood level and they ascribe that relationship to the subjects within the neighborhood. In other words, if neighborhood income is related to bad outcomes, then that means that at a personal level income is related to bad outcomes. Well that is not necessarily the case. It could be, but it doesn't have to be. Likewise here, if you were to look at this the aggregate effect is positive. As X goes up Y goes up. You can't ascribe that to the individual level. As a person's X value goes up Y goes down in fact. So it's similar in concept to that here applied in longitudinal. The good thing about it is that we can easily asses and test this assumption.

So I did that and at first I was a little worried because I had published results assuming that those effects were the same and I thought not again. Do I have to again send in a retraction of the results? Like on all my papers I got to send in an errata sheet or retraction or something like that.

Luckily when you go for tenor they don't look at those- No, I'm kidding. I thought no, what if I had bad results? Here on the top shows the model that I've estimated up to this point assuming that they're the same. Down below I've relaxed that assumption. Again, what I need to do is to create two versions of each of the variables. What I'm labeling as between subject or BS here. BS here stands for between subjects. It is the average of the drug level across the four time points and within subject is at a time point how is the person's value different than that average.

Now the first thing I want to take a look at is the deviates because we can do a likelihood ratio test to compare these two models. Say that's three on two degrees of freedom. This second model has four effects for the time variant covariate. The above one has only two. The difference is the two degrees of freedom, so I don't have to send a retraction in on these results. I can accept the null hypothesis and go with this simpler model that says the between and the within are in fact the same.

But if we look at the relaxing of it and explore that it's kind of interesting in and of itself. Look at DMI first of all. Here is the between subject effect. Here is the within subject. Again, the between subject effect the way you would interpret that is people's average level of Desipramine is related to their average Hamilton across time. Incidentally, this is change in Hamilton, what I used subsequently in the analysis, so there is a strong negative relationship. The within subject effect says relative to the subject's average, if they're higher or lower on the drug level are they higher or lower in terms of their outcome. Again, it's a negative effect. It's not significant here. Now you might make much of this, but realize in this dataset I've only got four time points and so going to this expanded model and separating out these effects yields these standard errors that are rather quite a bit larger than what they were in this model assuming that they're the same.

So what is important here is note that the direction is pretty similar and this averaging here makes sense because of this test result saying that the averaging that they're same is reasonable. Now for IMI though it's kind of interesting. Here we do get an example of them being of opposite signs. So it's like if you took their average of IMI across all time points and looked at that with relation to the average the Hamilton there is virtually no relationship. It's slightly negative, but not greatly so. If IMI is higher **** is high at a particular time point that would suggest that their Hamilton is high at that time point, so kind of a weird result for IMI and showing this changing of the signs.

I shouldn't make much of this latter model. Again, it's not in this case, worth it to go to this because we can accept the simpler model up above. Yes.

Question: Can I ask a question? For the within subject's variation you're assuming at a same one time point there is some variation within the subject for the DMI and IMI, but since you only measure- you only have one measurement at one time point the data, where do you get that data to testing the within subjects variation at the same time point when you have only one data point?

DON HEDEKER:No, I have four data points for both drug levels.

Question: Okay, but I mean at each time point you only have one measurement, so.

DON HEDEKER:So again, let's say this one. What this is, is the average of those four time points, the average drug level. Then what this represents is I take each one of these four I say okay, what is this minus this, what is this minus this, what is this minus this, what is that minus that.

Question: So but for the within subject variation should it be at the same time point, at one time point their **** measurement and then you can really test the within subject variation at the same time point?

DON HEDEKER:I'm not doing that here. What I'm saying within subjects is across the four time points.

Question: So across four time points.

DON HEDEKER:Yeah, within subjects across four time points. At a given time point if a person is higher or lower on drug relative to their average are they higher or lower in terms of the outcome? That is what that is getting at. It's not getting at within subjects within time point. No, it's not. It's across time points.

So this was kind of an interesting result. Luckily it was satisfied. Let me show you how you can do this in SAS, creating this two versions of the dataset or of the time variant covariate and the we can actually run this and then break for lunch. Now in this dataset I've created- I've got the ID. I've got the Hamilton depression level change from baseline. That is why I call it delt for the Greek letter delta, change, so that is Hamilton change from baseline, so if you look at these values they're largely going to be negative because they're higher Hamilton levels at baseline than at the four subsequent time points. Then I've got my old friend the intercept, which I don't need for SAS. Then I've got week, zero, one, two, three, four. Then I've got sex. I've got gender actually. I could look at if there is a different effect for males of females. We haven't explored that yet. Endogenous, zero or one, then I've got my IMI and DMI on the log scale.

Now to create- What I have to do here, notice in this dataset I don't have the mean of the time variant covariates. I need to get that. How I can do that is using PROC Means, so first I sort the data by ID. Actually the dataset that I read in is already sorted by ID, but I want to make sure of that, so I sorted it again, sort by ID and then I use PROC Means using the ID as a class variable and variables IMI and DMI on log scale creating a new dataset and calling the mean of IMI by ID MLN IMI, the mean of DMI by ID MLM DMI, so that is how I can create the aggregated means of these time variant covariates for each subject.

Now this dataset that I'm naming two has within it basically 66 observations, one for every subject and what it has in there in the dataset is their ID and it has the mean of the IMI, the mean of the DMI, so now what I got to do is merge this back together with the original dataset, so that is what I do here. I merge this created dataset that has the means with the original dataset I read in by ID into the dataset I named three and then I create the deviations of the time variant covariate relative to its average across time. I named those L, NI, deev [ph] and log D deev, so

that is how I've now created here the mean of the time variant covariates and their deviations relative to those means.

I now can do these two analyses, the first one where I assume that the effects are the same and the latter one where I relax that assumption by containing both versions of the time variant covariates and exactly the last slide I showed did a likelihood ratio test comparing these two. Found that this one did not fit statistically better and in that case I can use the simpler model that doesn't make this distinction between the two types of effects.

So let's try to run that. So we'll go back to SAS and let me read in that. So the dataset we want to read in is named RisBSWS [ph] for Risbee between subjects within subjects. This has exactly the syntax that we saw on that last slide. I believe it does. Again, I'm going to have to change that in file because the dataset we're reading in is here. Let's take a look at that dataset. It's RisbeeT4dat [ph], so I'm going to open that. So here is the dataset that I'm reading in and going to be analyzing. It has- first is the ID. Notice that there is four records per subject approximately. The value of the **** change score, Hamilton change score, so it's negative, the values of ones that I don't need, the values of time zero, one, two, three, then I think there is gender and endogenous. I'm not using those variables in this analysis. Then log IMI and log DMI, they're on log scale.

Does anybody notice anything weird in the dataset? If you do you get a trophy. I don't have it, but. Yes.

Question: There is no missing Hamilton.

DON HEDEKER:True, that is one thing. In this dataset what I've done- In the last one I had six observations per subject and I had missing values coded with a period and then in the analysis what it did was to throw out those ones with the period. Here instead what I've done is just to put in however many time points a subject has, so for example, subject 113 only contributes at 012. They don't have that final time point. The two approaches will give the same results. I'm physically throwing those observations out here rather than letting SAS throw them out at the analysis phase. That is what I'm doing. That is a good observation, but there is another one, another thing you might observe here. Any value of a particular variable that kind of stands out? What is that? Negatives, they're change scores relative to baseline.

Question: There is a zero for the ****.

DON HEDEKER:There is a zero, yeah. Zero for this **** log of DMI. I didn't discover that until about a year or so ago. And so how did I discover that? Actually it was with the program Super Mix. Super Mix has the ability to do what are called multivariate plots. Multivariate plots show- you read in a dataset and you get the bivariate plots of all the variables, so for example, you'll get a bivariate plot of Hamilton by DMI, Hamilton by IMI, DMI by IMI. You get all of the bivariate plots. You can do that actually now with SAS too come to think of it and when I did that I discovered this one weird observation on DMI. Now I should have discovered it when I looked at the dataset.

So what is that? It's a value of DMI of one **** transform the zero. Let's run it now and then let's run it without that observation and we'll see something kind of interesting that took me years to find out and you guys have figured it out in a matter of seconds. You're way ahead of me. So we'll go back to SAS. Let me get the name of this folder. All I've done is to change this directory, so everything should be okay. Let's run it and get the results of this analysis. I should have- Let me go back and clear out the output just because I had all that other stuff out there. Let me just clear that again, clear all and I'm going to clear the log file as well. Now I'll run it again. All right, it ran pretty fast.

So this first run, this is the one that assumes that the between and the within are the same. There is 66 subjects. Maximum number of observations is four because there is four time points. There is 251 observations used here. Incidentally, you might wonder about this number of observations used, one. Let me show you something. Let's go back to this file here. Let's see, PROC, print, data equals two because when I saw that I was like what is that all about. Let me show you this. I'm going to clear the output again. So all I'm doing here is I want to print out that aggregated dataset that I've created and I'll show you why in a second. What happened? I hate those semicolons. Let's try again. There we go.

Now when I ran this I assumed that when I did PROC Means it would create a file that had 66 records, one for every subject, but remember how assume spells. When you assume something- Let me show you. When you print it out you actually discover something. Look at the observations. It starts at 1 and ends with 67. There is actually 66 subjects in this dataset, but there is 67 observations in this dataset. Why? The first one, what the first one is, is an average of those variables across all of the observations. Now notice that it doesn't have ID present, so in my mixed analysis it's not used because it doesn't have an ID variable, but that is why you see when I did that run it says one observation not used. It's that one up there.

Now here is one thing and admittedly it's like why is SAS doing that? I asked for means of these two variables by ID. I didn't ask for that. That is the mean across all IDs, but when you do means that is what it does. It puts in this overall aggregated version in the dataset. Now if you don't like and admittedly I don't really like that, here is one way to get- that you can remove that one, so let me modify this in this way. I'll say data 2B, set 2 if ID not equal to period. Yes, question.

Question: Can you use **** option **** like you can in ****?

DON HEDEKER:You might be able to.

Question: And just **** the last.

DON HEDEKER:Yeah, there might be a way like that to do it too. Yeah, that's true. I don't know. I'm not sure, but yeah. But this is another way you could do it is you know what I'm doing is reading in- bringing in this dataset, creating a new one, but only if ID is valid, so it will get rid of that first observation, but as you mentioned there might another way to do it within PROC Means.

So let me run it now with this and let me change this merger here, the 2B. I think then I should be okay. Let's see, so I'm going to clear the log, clear the output and run this. Let's go to the top. Yeah, notice here number of observations read it's 250, number of observations used 200, so that average didn't get into the picture at all here. Again, I want to reassure you it wasn't in the analysis even before because it didn't have a valid ID, but still in your output you got this number of observations not used one and you might go to a person and they'll say what is the observation and then you got to go through all this explanation, it does this, so if you want to avoid all that just get rid of it the way I did there, so you can see the nice zero here.

So here the results of the first analysis treating the common, the within and between subject effect. here is the variance, covariates matrix of four, the intercept in time and then following this we get the- and let's jot down this deviance here, 1498.8 and this is a solution that I've mostly been focusing on for the effect of drug, highly significant effect of DMI, non-significant effect of IMI and let me jot down that effect of DMI. I want to jot that down, negative 1.97, so it's a pretty big effect. One unit change on the log of DMI results in a two point change on Hamilton and it's really comparable in size to the time effect in this case and then subsequently here is the effect separating out the within and the between.

Now here we get the deviance 1495.8, so again, you do that bit of subtraction you get the chi-square value of 3, so that was the chi-square test of within and between are the same, chi-square of 3, a 2 degrees of freedom, not significant and here is the separation of the between and the within that we looked at. Again, the effect of DMI you can numerically are pretty similar. The effect IMI not so similar, but the standard errors are rather large, so there is no statistical reason to prefer this model to the simpler model, which assumes that the between and the within subject effects are the same.

Now finally, let me show you what that one observation on DMI is doing here. So remember when we looked at the dataset you astutely pointed out that there is this weird observation on DMI and it's rather outlying actually because most of the values you can see are like four, five, three, so it's quite a bit separated from the bulk of the data on this covariate. Now I mentioned the Bill Gates thing. We're going to see this actually happen right now because our previous estimate for the DMI effect was negative 1.97. Now let's see. There is 250 observations here. Let's see what happens when we remove this 1 observation, 1 out of 250, so let me go back here to the SAS code and I'll put it in here. I'll say if- Sorry, wrong code. This one. I'll say if LN DMI is greater than zero, so what I'm doing when I'm reading in the original dataset I'm only permitting in the observations in which the log DMI is greater than zero, so in other words I'm going to exclude that one observation and then I'm going to carry out the rest of the analysis just like I did.

So now we should have only 249 observations in the analysis. So let me clear out the results just so that I don't get confused.

Question: Without **** clearing out if it happens that the value was not zero, but was .00000001.

DON HEDEKER:Yeah, you're right. It may not. Let me be a little bit more- Let's see if it clears it out, but you're right, maybe it won't. I think it will, but let's take a look at that data. It's .0000. Let's see if it clears it out. 249, so it got rid of it. So here is the analysis assuming that within and between are the same, which we've seen to be a reasonable assumption here, but taking out this one very low value of DMI, 249. Here is the random effect variances. Here is the estimates of the covariates. Look at the DMI effect. It was negative 1.97, negative 2.3. One observation out of 250 made .4 difference in the effect. It's more negative, more significant in this model, so yeah, that observation was really messing things up.

Now when I discovered this I was like is that a valid observation or not, so I called up Risbee and found out well this study was carried on in 1977 do you expect me to have the actual data? No, I didn't call Risbee. Now I couldn't go back to the raw data per se. This is the data I had, but it made me really suspect that particular observation because it was so- Why would they have DMI coded one on that when all the other values are like these big values? What was-? Maybe it was supposed to be 10. Maybe that was a missing value code. Why would it be an integer because it was an integer one and the drug levels were largely non integer values? So I don't know for sure, but I would suspect that was some kind of problem with that observation. That is not a correct data point I feel. Now I don't know for sure the answer at this point because the study was carried out so long ago and I didn't pick this error up until recently, so I would say this. Do as I say, not as I do. If I would have done a better job of looking at that data initially I probably would have discovered this much before. It still might have been too late, but having discovered it and looking at the results I'm kind of mortified in some ways that the effect changes that much.

Now luckily for me it's still negative and significant, so you can say so who cares, but that is a rather large difference in terms of the point estimate, negative 2.3 to 1.97. That's not a trivial difference, again, for one point out of 250, so it really taught me an important lesson too is how sensitive some of these things are to outlying values. Again, this was an outlying value not in terms of Hamilton per se, but in terms of X. Really again it's a real example, an **** example that the thing I described with the Bill Gates thing. You have an X value that is far off the beaten track. This wasn't far off the beaten track in the positive direction, but in the low direction. It's still far off the beaten track. It can have a great influence on the estimation of the slope.

Now what we're seeing here is that the slope is more negative when I take that out, so what does that suggest to you about what is the value of the Hamilton for that observation, for that one observation? It's not so negative in other words. Let's go back to the dataset and you'll see what I mean because basically by putting in that one observation it's making the slope less, not greater, so in other words its value of Y is not consistent with the rest of the observations and in fact, when you look at the data point you get a feel for that. Yeah, look, it has a Hamilton change score of two, two. It's a times zero. It also leads me to suspect it's at the first time point for this subject. It could be that maybe the subject started- Well they got IMI though. I was going to say maybe they didn't get the drug, but no, they get this. It's not like they take both of these things. This is a metabolite. I don't know. Maybe they couldn't measure it for whatever reason that first week. Maybe they forgot to measure the metabolite for that subject that first week. They only got the parent compound. I don't know, but it has a value of the Hamilton that is rather non negative and so that is making that slope less negative than if I take that out.

So again, looking at these outliers is kind of embarrassing when you look at them too late and truly in this analysis I think that this result without that one weird observation would be a more valid one, but I can only suspect that. I can't know for sure because I don't have the original data forms and can't really do anything to confirm or refute my suspicions. Question.

Question: But if you- say you have the dataset and that is really a true value and you measured it yourself would you throw that out as ****?

DON HEDEKER:Yeah, so that is a good question. The question is let's say you went back to the datasheets and it was in fact recorded as one, it's as far as you can tell a true value what should you do? What can you do in that case? Well in that case it's true. You would be obliged to keep that data point in the analysis really because taking it out I mean if it's a legitimate value it's a legitimate value. That is what you can do. Now what I was going to think about is like well maybe you can do some kind of transformation that would minimize its impact just like I took the log to minimize the impact of the higher values, but to be honest I'm trying to think what could you do, what kind of transformation could you do to minimize the effects of the higher as well as this one lower observation. It's kind of hard to think about, but presumably that is what you might try. I don't know exactly how to accomplish this right off the top of my head, but that is what you can potentially do is to think about some transformation of the variable that will bring the observation more in line with the others, but you would be obliged to keep that observation in if you did find out that in fact it was a true observation. All right, any other questions?

Well I think we'll break for lunch a little bit early, but I put you through a lot this morning. I know of that and so I expect that after lunch they'll be probably not so many of you around. What we're going to cover in the afternoon is we'll cover dichotomist longitudinal outcomes and then missing data, attrition and really get into how you might explore that angle in longitudinal studies.