Multilevel Models for Longitudinal Dichotomous Outcomes—2010 MCH EPI pre-conference training workshop

DON HEDEKER: Feel free to complete that. I think we have till 5:00 or something like that. Okay, well what we're going to talk about this afternoon is longitudinal analysis of dichotomous outcomes. And so yesterday we talked about clustered analysis of dichotomous and this morning we talked about longitudinal analysis of continuous outcomes, of course now we're going to longitudinal dichotomous.

Here's another one of my websites that I have that has some of the materials that we're going over today and so you can download the datasets, which I guess you have anyway, but some additional resources are there. And again, some of this material comes from a chapter I worked on a couple of years ago for this *Encyclopedia of Statistics and Statistics and Behavioral Science* and then also a chapter in *The Longitudinal Book.*

So, this longitudinal outcome that I've looked in a lot of times dichotomous on smoking, non-smoking, right; it's one or the other. But there's many examples of dichotomous outcomes in the various literatures. Here's another good one. ****, it's only a two-step program. AA is 12-step, right. So it's a little bit easier. All right.

So, as we learned about yesterday, logistic regression it's a model that relates explanatory variables, co-variants to a dichotomous dependent variable and therefore multi-level or mixed logistic regression is a model that relates covariates to a dichotomous dependent variable where observations are nested. Okay?

So yesterday we talked about clustered scenario, today we're going to be talking about longitudinal where they have repeated observations within subjects. Again, logistic is a common tool for analysis of these, but you can also do this as probit regression models and in some areas probits prefer to logistic mostly in genetics work and econometrics, but in biostat and epidemiology it seems like logistic really rules the day.

So let's start with the random effects, random intercept logistical regression model. So just like what we considered yesterday when we had clustered observations, subjects within say schools or schools. Here though again, what we have are repeated observations within subjects. So consider the model with P covariates for dichotomous response YIJ subject I at type point J. Again, J goes from 1 to NI, that means subjects can have different numbers of observations. And the numerator here, we've got the probability of the positive response and the denominator we have one minus that or in ably the probability of a zero response, so that ratio is the odds taking the log, that's the log odds or logent.

And we have XBeta, our usual logistic regression part augmented with a random subject effect. Okay, so simple random intercept model here for longitudinal. Again, I want to make reference to this threshold concept to help us understand better the meaning of the parameters, the Betas, the scaling. So again, just like yesterday, let's consider that there is an unobservable continuous Y that has a distribution. We would like to obtain that, but what we get is a manifest version that's just a dichotomous version, YIJ. And so again, I've used fonts to relate those two. And basically, the notion is that there's some unobservable threshold if a person is above this the emit a one response, if they're below it they emit a zero response and depending on that distribution will either lead to normal or probit regression or logistic regression. And again, this helps us understand for one, why the scale of the betas is numerically different between these two models; logistic and probit.

Again, here is that same example. How was your day, what was your level of satisfaction today? Satisfaction may be continuous, but we often emit a dichotomous response. So the notion of the threshold concept is not really a difficult one and one that we actually experience every day pretty much.

So writing the model in terms of the latent continuous variable, just as we did yesterday, okay, here we have two sources of variation for the latent why we have the Landom effect, which is normally distributed, we have the errors which is either a norm or logistically distributed depending on our regression model. Again, we often use the logistic regression model and in some literatures, this model's therefore referred to as the "normal" logistic because we have normal random effects and logistic errors.

So this underlying latent variable, useful way of thinking of the problem, helping us to understand aspects of the model but it's not an essential assumption, so don't feel like it's important that this has to be… no it helps us to understand the model doesn't have to be. It helps

MICROPHONE PROBLEM FOR ABOUT 1 min. 30 SEC—CUT OFF AUDIO

**** from this 1.8 a factor of 1.6 long in sociology suggests 1.7. So it's important to realize that this is an approximation, but it gives us a sense of why the logistic estimates are numerically larger on scale than the probit estimates.

Similarly as we talked about yesterday, the estimates from a random intercept model and from a fixed effects marginal model, they're not on the same numerical scale. Okay. And with a random intercept model they would differ by approximately this amount, the square root of the design effect. In the numerator here we have what's the variance under a random effect model and there are two sources of variance, the subject variance and the error variance. In the denominator it's what would it be under a fixed effects model. Just the error variance.

Scott Zeiger looked at this a little bit more closely and just like **** Long suggested a little bit down weighting of this ratio, he suggested a little bit of down weighting of this ratio, a factor of 15 divided by 16^2. So where did he get that? I'm not really sure, but realize what it is; again, it's a ratio that's a little bit less than one there that he's putting in. So in idea, it's similar to what's been done up here. Okay.

Okay, so again, we can write the model as a multi-level model here in longitudinal, we've got the observed response, we could either write that as the logent or the latent response. We have an intercept for a subject, the time – the factor of time. Those get modeled down below in terms of level two variables. And then the random intercepts models we only have the intercept variance for the subject, but for the more general mode that we're going to get to, we'll have random intercept and time ****. So this is exactly like the model we were considering this morning, for example when we had group as endogenous, non-endogenous and we conceptualized random intercept and then subsequently random trend models. So, really you could write the model out in the same way regardless.

Here's a picture of a random intercept model. And here I'm writing it in terms of the probability. So the Y scale here is probability of a response. So what is this equation here? This is a logistic response function. Remember yesterday, the logistic response function is one over, one plus the exponential of – sorry. One over one, plus the exponential of minus logent. That's it… minus Z, and so in parenthesis here, I have my model for the logent, for the log odds. Here in this situation where I have group time and group by time and a random intercept, in parenthesis, that's what the logent represents. Now again, in terms of a picture, what I've done is illustrate this model in terms of two groups that are in the middle, subjects around the group trends lines and have put one above, one below, one above, one below to give a flavor of this heterogeneity, of course there will be as many lines as there are subjects. But on average, this is the group trend for the… looks like for the females, or sorry, males. This is the group trend for the females.

Now the random intercept model, they're parallel lines or are they lines. These are curves, why are these curves? Because the YX is probability. Right? And so we have this non-linear response. The "S" shaped curve of the logistic response function. Okay?

Okay, so realize, again, as I mentioned yesterday, the model is not linear in terms of the probabilities. In terms of the probability, it's like an "S" shaped curve where at lower **** in toted zero, higher **** in toted one. Okay? We are kind of at the lower end of that probability, so we're not getting the additional "S" shape in this picture, but that's the flavor of it.

Here's the model in terms of logents, log odds. Lines. Okay? So in terms of log odds, we do have lines because the model here is a linear model. So again, you can think of the log logents, it's basically a mathematical transformation that transforms the probability into a linear model. Gives us a, what we like on the right-hand side because we know how to deal with this. Right? Multiple regression ideas here, we have "X" variables,

those are things we can handle. So the logistic regression models are a convenient one to model that dichotomous variable in this way. Okay?

Again, I want to emphasize the scaling of the logent, .5 probability corresponds to odds of one, corresponds to the log odds of zero. Okay, so any negative logent value you get that corresponds to probability of less than .5, any positive logent value you get corresponds to a probability of above a half. Okay? So a half is like the anchor point that corresponds to this line here of zero logent, probability of a half. Okay?

Well again, what we saw this morning was random intercept model often isn't enough in longitudinal. Yesterday was enough for clustered, for longitudinal, not as much. So here we'll entertain the possibility that subjects have their own intercept and their own trend, but of those come from the bi-variant normal distribution in which you have three parameters.

The variance of the intercept. How much heterogeneity is there when time equals zero. The variance of the slope. How different are these response rates across time? And then their covariance.

So this is exactly like what we were considering this morning for continuous, albeit now we have, in terms of our outcome variable, we have a dichotomous response.

Okay, so maybe you're feeling a little bit like this… Statisticians, we find humor where others don't see it. Right? Okay. You might feel more like this. I know they like to bore us with statistics during the playoffs, but this is ridiculous! But I ask you, when are there not playoffs going on? It seems like there is constantly playoffs of some type or another. Right? What was it, last summer I was watching the World Cup and the World Cup ended and then I turned my attention to Baseball, and my wife's like, don't you ever watch something besides sports? Yeah, I do, but it always seems like there's something going on.

So we're going to look at a different dataset now to look at this dichotomous response across time and we're also going to look at this dataset a little bit later, we'll look at the missing data aspect of it. It's kind of an interesting study. It's also from psychiatry, it's a dichotomous response which is basically how well or ill they were. The syndrome here is schizophrenia and not depression and this was a study funded by NIMH, National Institute of Mental Health, looked on treatment related changed and overall severity.

So the dependent variable, it was really like an ordinal variable. It went from one to seven. And to be honest, I've analyzed this outcome as continuous, as ordinal, and we'll show it now as dichotomous and it gets kind of similar results, so I use it as a kind of a teaching example 'cause it's a useful dataset that highlights some nice aspects. Here what we're going to do is to treat normal, borderline, mentally ill, mildly ill as being generally okay. We'll call that, well. And then four and above as being sick. Obviously we're losing some information by doing that collapsing. Okay, so I want to impart that.

It's not necessarily what I recommend, but it's a good sort of teaching dataset, so we're treating it in that way.

The groups are these: placebo and drug. And actually drug comprises three different drugs; Chlorpromazine, Fluphenazine, and ThioridazineIn the study, there were four arms to the study. They were randomizing people to placebo and then one you know, each of the three drugs. Now in doing the analysis, the drugs were very, very similar. So here for simplicity, I'm just combining them into one just to make it a kind of simpler presentation. But realize that's why you got three to one ratio of drug to placebo.

Now what made this study rather interesting was the measurement across time. The protocol called for people to be measured at x0, that's baseline, week one, week three, week six. Now I never… I always wondered why those time points. Why zero, one, three, six. I don't know why. But there, you can see that there are some scattered measurements at some of the other time points, two, four, and five. Now a lot, but some.

The other interesting thing about this that we're going to get into a little more detail is the completion ratio. That is, how many of the subjects made it to week six. In placebo group, it was about two-thirds in drug group, 80%. That's quite a … that seemed to me to be a big difference. You know, why are you getting more completion in drug group, and does that matter? To be honest when I was first doing these analyses and I didn't understand missing data aspect so much, I was always concerned about his difference.

So you might be a little concerned about this difference. This afternoon we'll learn that it's nothing that you should be overly concerned with. Let's say that. So you can maybe raise one eyebrow, but not both. Okay? It's not that bad. Okay.

Now the main question of interest here is really, is there differential improvement for drug relative to placebo? You know, treatment relative to control. And basically what this is getting at is, is there group by time interaction? Because there's two groups that started out basically the same, at time zero. One group gets placebo, one group gets drug. So is there differential rate of improvement across time for the drug group relative to the placebo? The treatment group well to the control?

Now let's look at some descriptive statistics. Here in terms of descriptive statistics, I'm gonna focus on the four dominant time points. I want to stress that I'm gonna use all of the data in the analysis. Okay? However, to get descriptive statistics it seemed to be at time points two, four, and five, it didn't make much sense because there are so few measurement at that time point to get the proportion of people that are sick at those time points, it's based on rather scarce data, so I just calculated the proportions at week zero, one, three and six ****, but I'll be using all of the data in the analysis.

Okay, so here at weeks zero, one, three and six, we have these observed proportions, these are the proportions of people being above… at or above moderately ill, so at week zero, almost everybody is at or above moderately ill because that's why they're in the study. Week one, it goes down; week three it goes down; week six it goes down, and it

seems to go down more for the drug group relative to the placebo group. So there seems to be a treatment effect here.

Okay, so these are proportions. Everybody can understand proportions. The next sort of paragraph I have is the odds. So, I'm sure most epidemiologists can understand odds better than I can. You take the proportion divided by one minus that proportion, that gives the odds of being sick. So at week zero, the odds of being sick are pretty extreme because everybody is more or less sick at week zero. But those odds of being ill diminish across time, again, diminished seen more so for the drug group relative to the placebo group. And so there seems to be some effect of drug. We're going to analyze that.

Now notice that the odds of being sick are all above one, except for this one observation at week six for the drug group. And that corresponds to the one proportion estimate that's below .5. Okay? So again, any proportion that's above .5 yields odds greater than one and yields positive logents. So when you get to the logent paragraph of things, all the logent estimates for the two groups are above zero with the exception of this one for the drug group at week six.

So being negative on the logent scale here is good, right, that means that your probability of being sick is below a half. And that's only this one observation for the drug group, odds less than one, proportion less than .5.

Now the other thing that I've calculated there is the odds ratio. Okay, at week zero, I'm taking the odds ratio of placebo to drug. And there it's actually below one, but I'll point out that the odds of being ill are rather extreme for both groups. And so the .65, the fact that it would seem like, oh being in the placebo group is protective for being sick at week zero. Yeah, but that's because base rates are almost close to one and the placebo group is a little bit less than one. So I wouldn't make too much of that.

At the subsequent weeks, the odds of being ill are two-fold of the placebo, four-fold for the placebo group, 3.4 for the placebo group. So conversely, drugs group has a greater odd of being well. If you take the logents and you take the difference and you exponentiate that, you get the exact same answer. Okay?

All right, so that's essentially another reason why when we get these logent estimates, we can exponentiate them when we get the odds ratio interpretation.

Well one thing that we'd like to do is to plot the summary statistics across time. This morning we would plot the means across time and this would help us figure out, well how do we want to treat time in our analysis? Now here are the proportions across time; zero to one, and again, I got this **** proportion of .5, solid group is the drug group, dash group is placebo group. Now, let me ask you; should this plot inform me in terms of how I should treat time in my model. In other words, what I'm contemplating here is do I need time, times squared, how do I want to model the change across time? Can I get away with just using a linear affect of time?

Well remember this. The model is not linear in terms of the probabilities. Right? So this plot is kind of useful, but it shouldn't – you should rely on it to heavily to help you figure out, well how should I treat time in my model. Right? Because the model is not a linear model in terms of probability; it's linear in terms of the logent. So to help you decide how you should treat time in your model, this plot is not that helpful, but rather, this plot is. The logent plot.

Now if you see that logent plot, what do you – does it look like there's a linear effect of time? Not so much, right? I mean, for both groups what seems to be happening is really, there's quite a large descent between zero and one. And then thereafter, then – so you can certainly postulate doing what's called a peace-wise linear model. Once slope for the time between zero and one and then a second slope thereafter. And that seems like that would work pretty well. That would require two variables for time and therefore two groups by time interaction terms.

Now when I was first analyzing this data, I was working with psychiatrists and we were looking at a first continuous scale, but the plot of the means looked kind of like that as well. And so I suggested, well why don't we use a model with a linear term and a quadratic term? Time and time squared. That'll get at sort of the curvature and the response across time.

Well, my psychiatrist, he didn't like that idea. He said, "No, no. I don't want that 'cause then we'll have two terms for group by time interaction. And I want…" And I said, "Well we can test for those degrees of – we can get a two degree of freedom test for group by time interaction." He said, "No, no. I want one degree of freedom for group by time interaction." So, I think like, well how am I gonna get – eventually what he was wanting me to do is to find a way to linearize a relationship between the dependent variable and time so that I would only have one term for time in my model and therefore one term for group by time in my model. Okay?

Now, it wasn't such a bad idea because that would give me a very efficient test of group by time rather than my thought of having linear and quadratic. So I played around a little bit with various transformations of the dependent variable "Y", and then subsequently the "X" variable, time, and I came upon this transformation, the square root of week. Okay? Now admittedly, it's kind of weird. Right? But then I would ask you, what's so special about week? You know, is there anything magical about week. Well easier to understand week than square root of week, true. I grant you that. But what this achieved was the – what the psychiatrist was asking me – wanted to do. I want one term in my model for time. So I can have one for group by time. So the model was simpler; the interpretation, not quite as simple.

So that's what I'm going to show today. I don't say that this is something that you will always want to use, but it's something to consider in your toolbox is, you know, if you want a simpler model, you can think about transformations, especially of time, to linearize the relationship, and then you don't have to do – you won't have as many parameters in your model for time, for group by time, and in this case what it's going to

achieve is, I'll have a one degree for freedom test of group by time interaction. The thing I'm most interested in.

So I'm gonna be using "X" time as the square root of week. That's what led to this. And again, it's got certain pros to it; it had certain cons to it. It's not necessarily the best answer, but it is one way to approach things. Again, it wasn't my idea. My idea – but I can kind of appreciate what my psychiatrist friend was getting me towards.

Okay, before we get to the mixed model, let's run a fixed effects regression model. It's ordinary logistic regression. Okay, so here's the ordinary logistic regression and I want to emphasize this point. This is not an okay model because the data are not cross sectional longitudinal. So I say that about, okay if data were cross sectional longitudinal, by that I mean if it's different subjects at the different time points. But that's not what this is. The subjects are the same subjects across time.

The other occasion which this might be okay would be if the subject variance or here, the standard is equal to zero. In other words there was no data clustering. I know both of these things are basically false, so I put this here just for curiosity and to kind of something that we're going to compare our subsequent models to. So here we have a fixed effects ordinary logistic regression analysis. Well, when you got these estimates of stated errors, why don't we take a look at the "P" values, just for the heck of it, okay. We won't write them in our journal article because we know that they're wrong, but we can't resist interpreting **** these "P" values. Okay.

So what this is is the drug effect. When you have drug by time in the model, this represents the drug difference when time equals zero. Okay, so that's the base – at baseline. There's no difference between the groups. Well, in a sense we knew that already because the subjects were, you know, that was before intervention began. So at time zero there really should be no difference.

The time effect. This is the time effect for the drug group coded at zero. Again, because I have the interaction in my model. So this is the placebo time track. It's negative, so the log odds of being sick diminish by about one point per square root of week. Right? Okay. Per time, let's say. And that's highly significant. So what's that saying is the placebo group is actually improving across time, based on this model.

Drug by time interaction. It's negative, which suggests the drug is improving at a faster rate, but it's not quite significant in this inappropriate model, okay. Because this model is not right because it is not taking into account the fact that these are repeated measures. Okay, so that's what you get out of the fixed effects. And we're going to use this for comparison purposes.

Now when you estimate a model, you might want to check, you know, what's that fancy model telling you, right. So let's check the results.

So here what I've done is to plot the observed logents for the two groups plus the four dominant time points, week zero, week one, week three, week six. And I estimated the trend lines for the two groups using these estimates from our logistic regression model and for drug plopping at zero or one depending on whether it's placebo or drug. The values of time, 0136, the square root of those values, and then drug by time interaction. And I get these two lines and I get the fit with the observed quantities and it all looks pretty good. You might squawk about this over here, but again I want to remind you that, really any logent value that's up in this range is corresponding to a probability of near one. So the fact that we're slightly below that is, it's kind of an over exaggerated point. The probability up in that region once you get like above three or below minus three, you're talking about a probability that's very close to one and/or zero. In fact, if you do this plot on the probability scale, then that goes away.

Now this is a logistic curve, but realize, it's not – it's in this direction because the probability is going down. So, earlier yesterday, we had the probability going up across time, here it's going down, so it's more like this. And here's the logistic response, **** again, and it shows that the model does a pretty good fitting of the observed quantities. Okay, so using the square root of week has saved us in the sense that we only have one time trend, one drug by time interaction. So we've got a rather simple model here to represent the proportion change across time.

Okay, well let's get to mixed model. That was for comparison purposes; now let's actually take into account the fact that the data are longitudinal. Again, I'll start with a random intercept model, just like what we did this morning.

Here are the estimates I get. First of all, down below is the likelihood ratio tests comparing this model to the simpler random intercept model… I'm sorry. This is the random intercept model; I'm comparing this to the fixed effects model. Okay, chi square is over a hundred one degree of freedom. Highly significant. Right. So what that says is the last model that didn't take into account the fact that these were longitudinal data, you know, is a garbage model. It's not a good model. This model fits statistically much, much better.

And her are the estimates that I get. These are logent estimates so the drug effect at time zero is very close to zero and highly non-significant. Again, some people have argued this. In a clinical trial like this, in which the subjects are really the same at baseline, there's no need to have a term like this in the model because we know the answer. In the population, the two groups are the same at time zero.

Now some others have argued, well this will then give you a test of whether, in fact, they are equivalent at time zero or not. So that's kind of a debatable point, but truly, in the population, they're the same at time zero. And if you do see a difference that's presumably due to sampling variation.

Now it doesn't matter here much at all because the point estimate is so close to zero and it's highly non-significant. I just raise this kind of just a little aside that's out there.

Now time effect. This is time effect for placebo group. Negative one-and-a-half, highly significant. So giving people placebo does help. Right? This is kind of known in psychiatry that there actually is a placebo effect by and large, right. Drug seems to work better, -1.0024. Okay. So this again, this is an example of when we did the fixed model, here I was getting paid only $10 an hour, right. When I went to this one, I converted that .11 to .0024 and now I'm getting $100 an hour. No, this is a more plausible analysis 'cause this is taking into account the longitudinal nature. We now are seeing a significant result of drug by time.

And then finally, we'll see a very strong measure of the intercept variance and expressed as an ICC we get .58. Think about yesterday when we had clustered data. The ICC for classroom was like about five percent, for school even less. You know, it wasn't .5, it was .05. Here it's like 58. Sixty percent of the observed variation is at the individual level. Individuals have a huge influence on their response across time. The data within subjects are highly correlated. That's why in longitudinal data, we need to do a little bit more work because this ICC is so strong; the data are so correlated that we really need to model it in a valid way.

Okay. Now this is a paradox kind of slide. I took these estimates and I wanted to do the same thing as I did for the fixed effects model. I wanted to answer the question, how well does my model fit the observed quantities. So here I've got the observed logents for placebo and drug… sorry, drug is the square, placebo is the open circle. And then I generated the estimated lines using these estimates from my random intercept model. Again, plopping in the covariate values of drug, zero or one. Time, 0136, square root of week, and then I go to this a random intercept. I said, well what should I put in for this? I know, I'll put in zero. It has mean zero, so just put in zero. And those the two lines I got.

Is there anything that looks weird about those two lines? First of all, do they fit… do they seem to fit the observed quantity? Not so much. Would you want to submit that plot to a journal article after you – no you wouldn't' want to do that, would you. You'd be like, something kind of amiss here. Right? But does it seem to capture the general pattern? More or less. You look at it, you see **** is kind of a scale difference or something. You know, there's something I haven't really taken into account.

Now remember when I made reference to that underlying latent "Y". And I made the point that the estimates from the mixed model are not on the same numerical model scale as a fixed effects model. I haven't taken that into account. These estimates here are what are called conditional. They're accounting for the random subject effect.

So using them to try to fit the marginal observed logents, just using that as-is, is no good. They're not on the same numerical scale. So that's one way to understand what's amiss here.

Now here's another way to understand what's amiss here. I put in zero here because it was the mean of the random effects. Let me ask you this. Let's say you got a series of numbers and you take the mean of those numbers. Now you take a series of numbers – you take those same numbers you **** transform, you take the mean of the log transform values. Does that mean there's a log transform values equal the log of the mean of the original values? No. Not in general it doesn't. Why? The logarithm is not a linear transformation. It's monotonic, meaning that as things go up, they go up, but it's an s-shaped curve. So mathematically, one way to also think about what's wrong here is, I put in the mean of the random effect here, that doesn't mean that I'm going to get the average logent, and that's what I'm trying to accomplish here. What's the average logent for the different group at the different time points? It doesn't work that way. It doesn't work that way. So there's two ways of potentially realizing there's something amiss here. One is that realization that the scale of these betas are not on the marginal scale, the fixed effect scale. The second way you can make rise to this mathematical argument, putting in the mean of the random effects does not give you the mean logent. Because the logent transformation is not a linear one.

That second point helps us to understand this. Yesterday, or this morning when we dealt with continuous, we don't' have to worry about that, right. There you have on the left-hand side of the equation all you've got is "Y" for dependent variable. There's no transformation. So plopping in the mean of the random effect there gives you the mean of "Y". Here it does not because we don't have "Y" there. We've got a non-linear transformation of "Y".

So in categorical data, I mentioned this yesterday too, but with clustered you don't see it quite as much because this effect isn't quite as large. Here you see it quite a lot. I mean, look at this again. Let's take a look at the estimates we got from the fixed effects model: 3.7, -.4, 1, -.4. Look at the same estimates we get from the random intercept model: 5, 0, 1 ½ , -1. By in large, these are pumped up numerically. Again, they're kind of like on steroids. Why? They're estimated adjusting for the random intercept in this model. They're not on the same numerical scale as the fixed effects estimates.

Well what can you do? Can you just throw up your hands and say, oh well? I guess in this article we won't use a graph. You know. With this report there's ****. If they ask for a graph, say, oh sorry, we don't know how to do that. No, we got to do a little bit more work. Okay, so let me talk about it.

So here's from the onion again. Maybe it's not so bad. I mean, gee. Look at – here the pop for the census, right. U.S. population at 13,000 right? Actually we really had lunch with somebody from Hawaii and they told me that the population of Hawaii is a million, so this isn't even close to Hawaii population, let alone the whole United States.

All right, well something we don't need to worry about it if the census doesn't worry about it, but maybe you've got a boss like this. Commrade, the Comisar of mathematics wants it to equal 29.86, and you get .671. All right, so we've got to make it work somehow.

Okay, let me describe more explicitly what we're doing here in this random intercept logistic model. Now in this model, we're modeling the logent for subject **** .J with covariates X, but also with random subject intercept. So every subject has its own propensity for response in this model. And the influence of X is determined controlling adjusting for holding constant the subject effect. The covariant structure or dependency of the repeated measures is explicitly modeled because we've got this random intercept in the model. In rise it gives rise to a compound symmetry structure that we're going to want to go beyond, but at the moment, let's stay with this rather simpler mixed model.

So here's the really more proper interpretation of these betas in this mixed logistic model. Beta knot is the log odd's response for a typical subject with X = 0 and random effect = 0. So again, it's not averaged across subject, it's for the average subject.

The betas for our "X" variable, they represent the log odds ration for response associated with unit changes in X for the same subject value of the random effect. That's why I the biostat literature, sometimes the mixed model estimates here for dichotomous outcomes are referred as subject specific. More precisely, they're just subject adjusted. Subject specific is a little bit hard, but that's what they're sometimes known as. It tells us how a subject's response probability depends on "X". And we have this term, sigma square epsilon that tells us about who different our subjects in our sample or in the population rather.

So some people would argue using a mixed model for a dichotomous outcome is most useful when the objective is to make inference about subjects rather than the population average and/or there is interest in the heterogeneity of subjects. Now you might be very interested in learning about the population rather than these subject specific estimates, so what we can do though is to "marginalize" them, to give us also the population averaged interpretation. So let's get to that.

Before we do, let me show you some other interesting plots. At least I think they're interesting. This shows the response profile for all of the subjects in the placebo group. So previously what I showed was the response profile for the average subject. But now I'm putting in all the estimates that I have for all subjects in the placebo group, the hundred or so placebo subjects and calculating for each of these 100 subjects what's their probability of being sick at the different time points? It's like this. And what you notice is that there's a lot of heterogeneity across the subject. There are some subjects who do rather poorly in placebo. But there's some subjects who do rather well. There's a wide range of response of probabilities. Why? The random subject effect is a big effect. Subjects have a lot of influence on their outcomes.

Likewise, if you look at the drug group probability curves, you see a similar story. There's a lot of range of possibilities here even with this random intercept model. Some subjects are doing quite well, some subjects not doing well. Drug doesn't work for everybody. When you look at the average trend lines, you get this idea that drug always

trumps placebo. Not for everybody. Then you get some idea that across subjects, there's a lot of variation.

So here you get some idea that across subjects there's a lot of variation. So here what instead what I have plotted is not just the average trend lines, but the individual trend lines. Now, could I submit these to the journal article? If I say, no, we won't – we just want the average trend lines, we don't want to know about the individuals, how they're changing. That's all very nice and good and all that, but we only want one plot, we don't want all these crazy plots.

Here's a plot I like, but again, it's probably hard to sell. What I'm doing here is sort of incorporating both the effects of drug time – drug by time and the individual effect for both groups using one sigma below and above. So that's what I put in her. I put it here, I've got a random effect and I put in one sigma. Okay, the variance was estimated to be over four, sigma is estimated to be about two. So this shows then the two trend lines for the drug group, this would be one plus sigma, one below sigma, similar thing for the placebo group.

Now, why do I say I like this? Well you know what this shows me is this shows me the effects of the individual as well as the drug and time on the response. I mean, there's clearly some drug patients who don't do well. There's clearly some placebo patients who do fine. There's a wide range. Look at the range. Just one sigma above and below, not even going on two sigma, one sigma. There's a wide range of heterogeneity in the response profiles. People have a big influence on their response across time. It's not just drug, it's not just time, it's subjects. Subjects can you know, have a huge effect.

Okay, and they say, that's all fine and dandy, I still one to get one line for placebo group, one line for drug group. Okay. Well we'll get there. But it's going to require some additional work. So, what do they say? No pain, no gain? In these two days, you've had a lot of pain, right? With very little gain. Here's some more pain. Okay.

We're gonna fit those marginal proportions but we have to marginalize the random effect estimates. So I'll do this in these four easy steps. Okay? First, I'll calculate what are the estimated logents – I'm sorry, I should say, what's estimated there is the latent YXBeta. Then I'll calculate the marginalization factor. Again, all that is, is the square root of the design effect. Here's the way we've been seeing it, the variance under cluster sampling, variance on the random sampling, the ratio of the square root of that. Sometime you see it written in this way where they do a little bit of algebra on this expression to simplify it. Again, the error of variance is one per probit or here a standard deviation, PI divided by square root of 3 for logistic and D, again, is the design effect of the sampling literature, and then we marginalize. Okay, so these are now the marginalized logents.

Once we do this division of the XBeta part by the standardization or marginalization factor, now we can apply the logistical response function to get our estimated marginalized logents.

So in producing that first plot where things seem to be off, the part I left out was two… two and three, really. And what I did was I calculated this one and then I applied the logistic response function. So I didn't do this marginalization, I didn't realize the betas were on different numerical scale. Now I know they are; now I know how to correct for that.

Okay, a couple of notes. Again, in practice **** suggests a slight correction to that just using the standard deviation, this 15/16 thing. Logistic is approximate, it relies on the cumulative **** in approximation to the logistic **** here means normal. For multiple random effects, we're gonna get there in a moment. We have to expand our notion of what's the marginalization because there we have more than one random subject effect, but the idea is really the same. Will there get a marginalization vector depending on the random effect variance, covariance structure that we're having in our particular model. And then we perform element-wise division.

So a little bit more work but the concept is really similar. All right, now I did that with these data. Here's what I got. Again, on top you see the logent and you might squawk a little bit about that apparent misfit of time zero again, I would remind you that that's not much misfit in terms of the probability and when you look at the probability scale for the two, you finally say, hoo, now I can submit this to the journal. Right? Now I have a good model fit for the two groups across time, a nice expression of how the drug yields lower probability of being sick across time relative to the placebo group. Okay, so a little bit of a pain, but you get a little gain out of this.

Okay, let me show you the code to do some of this with SAS. Now in this dataset, it's a dataset that created that has I think seven records per subject and it has these variables. ID, again, we got to have ID. And then I've got three versions of the dependent variable; N79 is the original metric going from, I think on to seven. Then the binary version of it that I'm using right now, then an ordinal version of it that groups it into four categories. Then my friend, the intercept, which I don't need. Week – the square root of week and then treatment group times square root of week. On and then, sorry, I forgot about treatment group.

And then here's the coding of some of these variables: 79B that I'm using, zero is less than or equal to the mild, one is greater than or equal to moderate, treatment group zero for placebo, one for drug.

Okay, that fixed effect regression model that we saw, here's how we an estimate it simply in SAS. You can prologistic just like yesterday for cluster, we have this, need the descending option so that we're modeling the probability here of the one response and then via Glmmex, we can do the mixed model. Here's the random intercept ****. So it looks just like, basically, what we saw yesterday for a cluster, only addition is, again, we've got method quadrature 21 points and no CL print will not print out all of the different ID values. Class ID, the model statement… oh incidentally, yesterday I think I had descending over here. You can shorten that to DESC I guess in SAS. So we're

modeling the binary outcome with treatment, square week, treatment by square root of week. Solution: distribution equals binary, link equals logent.

I think actually link equals logent, you may not need it, it might be the default. Random intercept subject equals ID. Okay, so this is exactly how you can estimate the random intercept model using Glimmex.

And then, for those hardy souls out there how to do it with NLmixed. NLmixed, again, you've got to write out the probability statements. You've got to name the parameters and here I'm back to this convention, B0, B1, B2, B3, you know, again, not exactly the most information labels for the parameters. But it can work. Now, here's a good questions, here I'm calculating the ICC. I have an estimate statement to calculate the ICC. And in the numerator I have the variance of the random effects, and in the denominator I have the variance of the random effects plus this weird thing here. Can anybody tell me, that's for the arctangent of one times four? Does anybody have an idea of what the arctangent of one times four is equal to? Guess. Pi, right… it's equal to Pi.

Actually I didn't know that, didn't remember that. I would always go to my calculator and say, what's Pi, 3.1415, blah, blah… I would type that in. Alright, and I thought that we good enough. You know, like I get Pi to the eight digits. But then I saw somebody else's code and they had this arctangent 1x4, and I thought, oh, what's that? So I did some Googling I got **** some math sites and I found, Oh, that's Pi. That's what it is. So okay. So if you want Pi exactly, or more exact, it can never be – it's an irrational number. It's keeps going and going and going and going. But this expression, arctangent 1x4, that' Pi. So we have Pi^2 divided by 3.

Okay. Now, let's do the marginalization. Okay, so again, the marginalization involves these four steps. Here on slide 27. What I want to do is show it to you thought in a way that I can use when I get to models with more random effects. On slide 27, I'll show what you gotta do if you have just a random intercept model. In a random intercept model, the variance is the same across every single time point. It's just the sum of the subject variance and the error variance. In a random intercept and time model that we're going to be getting to, the variance of the time variable changes across time. What that means is that this marginalization thing is no longer the same value at every time point. It changes.

It looks like this. Now, notice the form on this marginalization vector. It's the diagonal of the estimated variance of the random intercept model, or random effects model, divided by the standard deviation in a fixed effect model. So take a look at that equation and the previous equation; square root of the variance of random intercept model relative to model without a random intercept. It's the same idea. Variance under the random intercept model… random effects model, square root of that, divided by the square root – the standard deviation under a fixed effects model.

Okay, so this form, the previous slide is a special case of this more general form. This more general form is useful not matter what kind of random effects model I have.

Random intercept model, random intercept in time, etc. All I got to do is work out this, what's the estimated variance of "Y", following this equation, get that, plug that into here and proceed. Okay?

So I'm gonna do that using this equation. This equation will serve me better as I get to other models besides the random intercept model. So here is how to do this with SAS PROC IML. Now, how many of you have explored SAS IML? All right, we have a few brave souls. If you've done any kind of programming in your life, you'll love IML. If you haven't done any programming in your life, you'll hate IML. I think that's basically the difference. Why will you love it if you've done programming? You can write things like Beta equals X transpose X inverts, X transpose Y, and it will do all of those matrix operations for you on one line. Okay, so they have all of the matrix things, multiplication, in verse, transpose all of those things. They've got those things built in. So it's really, to my mind, it's like a matrix calculator. So I don't have to deal with those thing at all, it will do all of that work for me.

Here we have to deal with matrices for this reason. We have to do this marginalization and deal with, what is the variance covariance structure of "Y". It's involving matrix operations. So we have to do this bit of work and IML is a convenient way to do it. So let me walk you through what I have here.

So I've got here the results from the NLmixed analysis of the random intercept model. And actually they're exactly the same as the Glimmex estimates as well. I'm first setting up covariate matrices for placebo and drug groups. X0 is the covariant matrix for the placebo group. It has four columns corresponding to intercept, drug, time, drug by time. Those are the four columns. The intercept drug; it's 0 for placebo group, time I'm putting her the four dominant time points, week 0, week 1, square root of week 3, square root of week 6. Drug by time interaction. The rows correspond to the four time points. So four columns correspond to the four variables, including the intercept. The rows correspond to the four time points. The comma designates that we go to a new row. Okay, so that's how it knows that we're going to a new row. It's not the fact that these values are a new physical line; it's the comma that tells SAS to go to a new row.

Okay, so this is a four by four covariate matrix for the placebo group. I likewise type in a four by four for the drug group. Again, intercept group, time, group by time interaction. I then type in my beta estimates for intercept, drug, time, drug by time, my variance, my random effect.

Okay, now I'm ready to do this marginalization of the random effect estimates, or the person specific estimates. Pi is equal to four times the arctangent of 1. All right. I'll always remember that. Well maybe not always. Number NT's for number of time points, there are four time points here. "J" function in IML. What that does is initialize. It's initializing a vector that I'm naming IVec. I'm going to be using this in a moment. It says there's four rows by one column with values of one. Likewise, ZVec. My random effect designed matrix here is four rows, one column, value of one. So the "J" is a way of initializing certain things.

EVec for error Vec. Pi^2 divided by three, again, I'm using this slight correction that **** recommend $15/16^2$ multiplied by IVec. IVec is four by one. So this has now got the error vector at four time points. It's basically Pi^2 divided by three at the four time points.

I then create my error matrix, the diagonal takes that Vector, which is four by one, and places it within a four by four matrix. Why am I doing that? Again, what I'm doing right here is precisely what's on this slide, this equation. Pi^2 divided by three times an identity matrix yields in this case a four by four matrix with Pi^2 divided by three on the diagonal. I'm gonna need to add that to this to get my variance of "Y", that I'm gonna then use in this manner. So I'm doing exactly what on slide 28 here in this IML code.

So we got through this, now I can create my estimated variance of "Y", it's "Z" times the variance of the random effect times the transpose of "Z" plus the error matrix. I then do this marginalization vector taking the diagonal entries, square root of that, divided by error, I now can marginalize XBeta for the control group and for the drug group, apply my logistic response function, print out the marginalized probabilities. Okay.

So that's what I needed to do to produce the estimates that I then used in this plot right here. For the logents for the probabilities. Okay?

So the pain is increasing every moment, right? Okay, let's up the pain. This morning we learned that random intercepts modeling for longitudinal data, it's not so great. It assumes that the variances are the same across time that all the co-variances are the same, the compound symmetry structure. So for continuous data, what we did was to add in a random time effect. We're gonna do the same ting here for binary.

So now, let's contemplate a model that has a random intercept and a random time effect. Now we've got two random effects just like we had this morning that come from a bi-variant, normal distribution.

So this is a very much akin to the model we saw this morning, but we had endogenous, non-endogenous that we had random intercept and time effect. Okay, now here are the estimates I get from this random intercept and trend model of this dichotomous outcome. Okay, the intercept that represents the logent when the co-variance equals zero and the random effect equal zero, that's positive, so and highly significant so what that's saying is the placebo group at times zero has a great log odds of being sick. Everybody's sick at times zero. So that's nothing new.

Drug. That's the drug difference when time equal zero, **** that's very close to zero and non-significant at times zero the two groups are essentially the same. Time is the effect of the placebo group highly significant, negative. Placebo group is improving across time. Drug by time interaction, negative and highly significant, drugs improving at a faster rate. Okay we're getting exactly the same conclusions as the random intercept model albeit this is a more sensible approach. How do we know this? Here's the test of this model relative to the random intercept model, chi square of about 22 on 2 degrees of

freedom, so this fits statistically better. And I would think, by this point, you will agree that it makes more sense to have a random intercept in time model because the random intercept model says variance is the same across time. Co-Variances are all the same. This is relaxing those strenuous assumptions.

Now here's the variance, terms, the intercept, the time, the co-variance, again, we get a negative estimate there. So the intercept and the time are negative and we associate it to some degree. There's significant variance in terms of how much – how severe people are at the start. Their trend lines. So it's exactly what we saw in the continuous case.

Now here are the estimated individual curves, in this case with the random intercept and time model for placebo group. There's much more heterogeneity because there's differences in terms the initial point as well as the curves across time. The "Y" axis here is probability, so these are like s-shaped curves going down, but not all. I mean, here's a person that seems to have started off with you know, middling probability of being ill and seems to have gotten worse actually across time, be even in placebo group there's some people who are doing quite well at the end of the study with good probability, very low probability of being sick.

Here is the two random effects that we put into the model to generate those plots. Here's the same thing for the drug group, again, the drug group is more dense because there's about 300 subjects there. And again, we see a wide heterogeneity in the responses across time. A lot of subjects doing well, but no everybody. So the drug doesn't work for everybody and there's a wide heterogeneity in starting points and slopes.

Okay, the here are the marginalized estimates for this random intercept and time model, logent, probability, so it's mapping on quite well to the observed probabilities, but after we do this marginalization. Again, that's kind of painful, no doubt about it.

Here's how to run the random intercept and time model. I first illustrated with Glimmex, and then I illustrate it with NLmix, again, these two yield identical results so there's no need here to use NLmix, per se. Glimmex here, I'm saying quadras are .11. Why did I diminish that from 21? Think about this, with one random effect, we're integrating over a standard normal distribution. With two random effects, we're integrating over a bi-variant standard normal distribution. It's bi-variant.

Here's what's called the curse of dimensionality. Does anybody know what the curse of dimensionality is? It's great. It's probably more fun around Halloween, but nonetheless, the curse of dimensionality. When you're integrating over a bi-variant, and you choose, say 11 points, as I've done here. You're using 11 points in both dimensions. So the total number of points that you're evaluating the data on is 11 squared. What is that, 121? Okay. So for each piece of data at each iteration you're making 121 numerical evaluations. That's a lot. And really take more time. So if I were to have chose 20 there, it would have been 400. It would be four times a slow, basically.

So, the issue is this. When you get to multiple random effects, you can't – you don't want to use as many quadrature points per se because it's really going to up the computation time. Now, luckily, both Glimmex and NLmix does a very smart method of quadrature know as adaptive quadrature. What it does it adapt the points to each person's location in that space. What that means, practically speaking, is that you can get away with fewer points. So, you know, using less point these days is not the issue it was 20 years ago. So 11 per dimension is quite all right, and so nothing to really fear, even though we may fear the curse of dimensionality. Okay, so that's what that is all about.

Then we have our typical model statement just like what we had before, here I got two random effects specified on the random statement, intercept and square root of week. Subject is people's ID, type equals unstructured. Again, I want to emphasize, in prox mixed or Glimmix, whenever you have a random statement and you have more than one random effect, use type equals unstructured because otherwise you're going to be assuming that the random effects are uncorrelated. That there's no correlation between the person's starting point and their trend. There is no A Prior reason to assume that. So you don't want that by default. Type equals unstructured says: estimate the full co-variant structure between the random effects.

G-Cor will then express as a correlation solution. Now what I've added here is a solution and what I'm doing with these ODS is to output the random effect estimates to this output location. So again, those plots I created with those individual trend lines, how did I do that? I had to do a little bit of work using the results in this data file. Okay? These have within it now, each person's random intercept, each person's random trend.

And here's the analogous code under NLmixed. Okay, we won't get into that too much. But you can see that it's similar to what we've had before where you have to actually write out the probability statements and give starting values and name coefficients and all of that.

Now here's the marginalization that we can do for the random intercept and trend model. The reason I made it so complicated before was that now the code that we used can now generalize here very easily. So again, what I'm doing here is to specify what is the co-variant matrices for placebo and drug group. So the columns correspond to intercept group, group by time, the rows correspond to the four dominant time points, week zero, week one, week three, week six. I type in what are the estimates of the betas for my four regessors, the variance, co-variance matrix of the random effects. I generalize PI, generate PI, say there's four time points. I create this IVec that I'm going to be using in a moment that's four by one **** of one. I define by random effect design matrix a column of ones for the intercept, values of week for the four time points, week zero, one, three, and six. And now I can use exactly the same code that I used previously to get my marginalized estimates.

So this is exactly like what we had before to produce the marginalized estimates for the placebo group and the drug group. And that's what I used precisely to give the estimates that I then created these plots with.

Okay. Well, let me say before we take a break, I want to say a little bit about GEE. How many of you know about GEE. Okay we have a few. GEE is an alternative procedure for longitudinal data analysis. What does it stand for? Generalized Estimating Equations. Very good. It was developed by a couple of statisticians, bio-statisticians at Hopkins. There's certain aspects of it that you're really going to love. One is that it's a "marginal model", what that means is, the mixed model I had to do all that marginalization. Here I don't have to do any of that. Okay? So that's good news, right. In fact the Beta here are on the same numerical scale like a fixed effects model.

How it counts for the correlation of the data is this. It says, well lets specific what's called the working correlation structure. Exchangeable in a one banded unstructured. And what it produces are estimates for the betas and their standard errors that are "robust" to model misspecification of the correlational structure. What that means is, you can specify the wrong structure and get tests that are "robust." Now the price – so that's good. Right? Obviously, that's good. The price you pay is, they may not be as efficient as if you get the structure right, and/or if you do a mixed model. What that means, less statistical power. So you're kind of buying robustness with some loss in statistical power. That's kind of the trade off here.

Somewhat of an easier model and more robust to, you know, getting the co-variant structure wrong, but the price you pay for this, you don't have as much statistical power for testing your various model parameters.

Now, it doesn't include any random effects, so it doesn't focus on heterogeneity, so as a result of that, the betas are on the population average scale. Beta knot represents what it would be like in an ordinary logistical regression. The log odds are responds among sub population with "X" equals zero. Beta, the log odds ration response associated with unit changes in "X" in a population of subjects. Exponentiating the betas we get our odds ratio that are referred to as population average. That's all great. And I would say, by and large, in the epi world, I can understand why GEE is probably a more favorable approach for binary than the mixed model is.

One thing I should point out, this is kind of curious, I think. For binary outcomes, GEE is probably more popular than mixed models. On the other hand, for continuous outcomes, it seems like mixed models are more popular than GEE in their application. But really, GEE can handle continuous or binary, mixed models can handle continuous or binary, it doesn't seem to me there is a strong compelling reason why for binary people will go with GEE and for continuous they go for mixed models. You know, it's like if you like the model, why don't you apply it for the other kind of outcome? You know, if you like GEE, why don't you use it for continuous ****? I don't know. That's kind of – it's kind of weird the way it's developed that way.

Okay. But for binary, clearly GEE has some nice benefits and easier to apply than all of the pain I've put you through this afternoon. Okay.

But let me show you this. Kind of curious result. So I applied GEE to this example, the schizophrenic study. And at first I used an exchangeable correlation structure. I said, well, you know, it shouldn't matter; the results are robust of getting that structure wrong. Here are my estimates. Okay, drug, no different at baseline. I've been seeing that. Time effect, that's the time effect for placebo group, negative and significant. Okay, there's a placebo response. Drub by time, negative and .1! Point 1! What are you talking about .1. I got .001 or something like that with all the different models I fit. Point 1? Non-significant drug by time interaction. I was thinking and perplexed by that ****. So then I tried some of the different structures. I thought maybe I just, you know, it's such an inefficient analysis. In other words, less power because I've got this work in correlation structure wrong.

That didn't quite solve the problem. I realized for this particular dataset, GEE wasn't really a good choice. Now, I want to emphasize that the particulars of this dataset are not ones that might be so common, so I don't want to knock GEE in a general way, but for this particular application, there are several aspects that made GEE less of a – not a good choice. Here's what those things were.

The working correlation, it's based on seven time points, weeks zero to six. What it's actually trying to do is figure out what's the correlation of that seven by seven matrix. Seven repeated observations. Well realize this. Several weeks have very little data and week zero is a near constant. Right? There's very little data at these weeks. How do you get a correlation between week four and five? There might not even be any person that measured at both week four and five. I mean, in this particular study, that's what happened. And week zero, almost everybody is sick at week zero. So what's the correlation of a variable with a constant? It's zero. It doesn't vary, a constant doesn't vary. Right?

So when I look at the working correlation, you know, the first thing I did was to jump to the estimates, the "P" values. And I was getting like, you know, scratching my head. I've done this too much. And then I looked at the working correlation matrix and I'm like, wooo, that looks weird. What's going on here? It made me realize that for this particular dataset the working correlation matrix is very poorly estimated and that seemed to have some bad effect on the standard errors and thus, the "P" values that I got from this analysis. Okay?

So if even you try analysis of four time points and unstructured, you know, the best you can get is to .05 with this approach. Okay, so with this dataset at least, it didn't work so well. Now again, I don't want to knock it in general, but for this dataset not so well.

Though, here's the results of taking those estimates and just you know, fitting the data that's quite easy. You don't have to do this whole IML, marginalization business that you people have so fondly considered. Admittedly, it's not so easy. Here's how to run GEE.

So here's a GEE analysis, you can do it in SAS using Gen Mod. And so here I'm running a GEE with just a regression model of this dataset and here what I did was to only allow the time points in which there was a fair amount of data, that is week zero, one, three, and six. I probably should have taken out zero too because as I mentioned, that was a near constant really for these data. And the model statement is just like what we've seen with various versions of the SAS, treatment times drug by time, link equals logent, **** equals binary. And then the repeated statement. That's what switches this to GEE because Gen Mod, when it first came out, it was a procedure to generalize linear modeling. So it can handle that. By use of repeated, it also does GEE modeling. So that's the switch that makes this more than an ordinary logistic regression, but a GEE logistic regression. There the syntax is a little bit different. Here subject equals ID. We've been seeing that on the right hand side of the slash in the mixed procedures, for whatever reason in GEE they put it on the left hand side of the slash. Within equals week. That tells it how are these data organized across time. CW prints out the work in correlation matrix; I would urge you to print that out because if you don't print that out, like I didn't print it out to begin with, you get some results that you might find kind of weird. When you look at this that, ah, give you a little bit of a clue about that.

And here is type equals unstructured that I selected here. Again, the four dominant types in GEE are what's called exchangeable, it's like a comp out symmetry structure. AR 1 and independent, that's like a **** structure, banded structure, and then unstructured. And so you can see how easy it is to run. Quite a bit easier than mixed model.

Okay. So let me conclude a little bit about what I've talked about this afternoon. I'd say that the mixed effects, logistic regression models are useful for incomplete longitudinal dichotomous data. It can handle subjects measured incompletely or at different time points. The missing data are assumed missing at random. What does that mean? Does anybody know the answer? Well, in about half an hour, you're gonna learn about the answer to that. 'Cause we're gonna talk explicitly about missing data in the final segment of this two-day onslaught of statistics.

So for the moment, we'll say missing at random, if you come back in half an hour, if you're not missing, then we'll learn about what that means. Okay?

Now here's one thing that you also need to consider when you have a dichotomous measure across time. The degree within subject variation on the dichotomous outcome is important to consider. Where this kind of bit me was, I wrote some programs to do mixed modeling for **** outcomes for a **** called Nixor. And then I put that on the web and made it available and then somebody was using it for an alcohol study and I think three time points binary outcome. Is the person drinking or not? And she ran my program, I think a random intercept model, and it wasn't converging to a solution, sent me this email, you know, what's going wrong with our dataset? What's with the program?

I sent back an email, how much did you pay for the program? She said, zero. What did you expect? No… send me your data, I'll take a look. Right. Hey, it was free, what do you want. You want something that works for free? Come on really? You thought it would work? No, it worked on simulated data. It works fine on simulated data, but real data?

No, I took a look at her dataset. It was interesting. It was binary outcome at, I think, three or four time points, actually. But the thing I discovered was, there was a large proportion of subjects that had either zero, zero, zero, zero, all zeroes or all ones. Now, think about this. If everybody has all zeroes or all ones, you have essentially no longitudinal information. You have no separation of within subject variance from between subject variance. Right? There's no way to separate those things if everybody is a constant across time. Now, in her dataset, it wasn't everybody, but a large proportion were like that. So computationally, what was trying to be estimated was very difficult because it was so longitudinal information.

Now I raise this point here with binary outcomes because you don't usually see it with continuous outcomes or **** 'cause there's more levels of the variable. You know, even like a continuous outcome across time that's relatively stable, you'll see some variation presumably in the scores. But with a binary with only two values across time, this -- you know, this can happen and it can kind of limit how far you can go in your longitudinal modeling if you don't have many subjects that actually change across time. Okay? So it's an easy thing to kind o miss, but also easy to investigate and so I would say, take a look at that when you have a binary outcome across time, see how many people are actually changing.

Now in the mixed model, we have this subject specific or conditional interpretation of the regression coefficience. We can do this marginalization, but it does require a bit of work, there's no doubt about that. Though I've talked about binary outcomes yesterday and today, there are generalizations to other categorical outcomes or no outcomes in particular, nominal outcomes as well when you have more than two categories, but they are not necessarily in ordered categories. And if you're interested in this, there are some materials that we didn't get to yesterday that talks about the ordinal models. Also on my websites, I have examples really of all these kinds of different outcomes and SAS code to do these kinds of analyses.

So, I'm sorry I didn't' get to it today, the ordinal outcomes, but if you're interested, again, there are materials there.

Well, we're gonna take a break in just a few, I don't know, maybe we should take a break now, it seems like that might be a good idea. And why don't we reconvene in about 15 minutes and then we'll learn about the mystery of missing data. Okay, so we gotta save the best for last. This is really the best. And I promise you, it's not going to be as complicated as what we just went over. I don't' think. It's probably going to be more complicated. No – it won't be that bad really, really. I promise.