ACTIV
ACTIV Cross-Trial Stats Team Meeting, Meeting Minutes

Friday, September 3rd, 2021

Action Items

Action Items		
	Information	**POC**
	No action items	

Notes

- **Frank Harrell:** [PRESENTING] This is our place where we put a whole lot of resources developed over the last year and a half. That address is hbiostat.org/proj/covid19. That'll get you links to all kind of things related to what I'm talking about today. What I'm going to go through now is the first bullet point there that has "new" by it. I'm going to give an overview of the Markov discrete time longitudinal ordinal model.
- I want to give some setup. What are our analytical and study design goals? And what's the setting? The setting is that you have a study where you have daily, weekly, monthly, yearly outcomes assessed. These outcomes can be binary, ordinal, or continuous. We're talking about a pretty general situation. Our ordinal model just reduces to a binary logistic model in the binary outcome case. So: what are your goals? To understand the time course of treatment, avoid picking a single day, such as a Day 14 ordinal outcome. Above all, we want to stay close to the raw data. We want to avoid data reduction. If you do time-until-event and what causes the event is not purely a binary process, you're engaged in some kind of data reduction that may or may not lose information. We're trying to stay as close to the raw data as we can—with one exception: if you have multiple outcomes on a single day, we would use the worst outcome that happened that day, although you can make a new category if co-occurrences are frequent. Let's say you have a stroke and heart attack on the same day in a large number of people. You can make stroke/heart attack be a combination that's worse than either one alone. It will handle that.
- We're trying to increase power by considering close calls, instead of reducing the data to time-to-specific-conditions. If you were close to triggering a designation of success or recovery, how close were you? For example, if you get statistical information from a blood pressure study, you don't analyze whether a given blood pressure was below a target, you find out how close to the target it was or how much better than that target it was. We don't want to have to decide whether an early milder event is worse than a later major event. We're not doing a win ratio global ranking approach that may require such a decision. We're allowing for complex correlation structure, including absorbing states. Our goal is to maximize power so you can have a smaller sample size or make earlier decisions in a sequential trial.
- We also want to handle missing data. If you're doing time-to-event analysis and the trigger for the event is using some assessments that may not be made on a given day, it makes it really hard to think about the time-until-the-ultimate-triggers hit—whereas with longitudinal analysis, because it's closer to the raw data, you can handle this fairly well with the standard missing-at-random assumption as well as using partial information. You might have an ordinal scale and one range within that scale may be a quality of life assessment that you don't assess on a given day, either by design or by a patient not being able to be interviewed, and so if you have one range of this scale that is missing, and if there's no clinical override that day, you would treat the unassessed scale

interval as interval censored. So you have a very formal statistical way to handle that kind of personal information. That just goes into your likelihood function for the probability of being in that interval, instead of being at a point that's uncensored. If you have a clinical event override, the missing assessments are irrelevant, because we would normally assume a clinical event is going to be worse than the scale you would have assessed on that day.

- So here's some examples of the degrees of power increases you might be able to get by using more raw data. The VIOLET trial is an amazing resource—it was a vitamin D randomized trial in critically ill adults, where the major thing measured was acute respiratory distress. This is from the PETAL network and NHLBI. The VIOLET study had usually high-quality data. It had daily assessments on a four-level ordinal scale every day for 28 days. The ordered scale is home, hospitalized, being on a ventilator, being in acute respiratory distress, or being dead. You can go to this link for a tremendous amount of detail and code and a lot of output. This handout here goes through simulations using the VIOLET study as the basis for fitting a model to simulate from. If your transition odds ratio is 0.8 in simulation model, the power of the Cox or log rank test for time-to-recovery is 0.2. But the power of the Markov proportional odds model using all the daily assessments is 0.9, so we're not talking about trivial differences here.
- Another way to look at efficiency gain is what happens when you use more days of assessment. So you have a choice of using a single day—let's say we use a middle day such as Day 14—or if we're using three days of assessment—say, Days 1, 14, and 28. If we're using five days, we might use Days 1, 7, 14, 21, 28. So if you filled in with more and more days, what is your effective sample size per subject? What this shows is if, for example, you use Day 14 as your outcome, your sample size per subject is one, so each, each subject is contributing one observation. But if you made your number of assessments go closer and closer to daily—this would be 28 assessments—you can see that it's more than a factor of four that your effective sample size increases. In other words, if you had 28 daily measurements, you know that there's really high correlations between those days, but even with that, having 28 days is like having seven patients randomized. That is quite a lot of increase.
- So what are advantages of discrete-time Markov models where you have one method for binary ordinal nominal—because you could have a polytomous logistic regression as the heart of this—or continuous outcomes? You have a general correlation structure allowed for including nearly absorbing states. Patients in the hospital for many days tend to change lists from day to day. When somebody gets put on a ventilator, they might stay on the ventilator for many, many days in a row. In other words, patients in the hospital often get more stable the longer they're in the hospital. How would you account for that? Well, easy because that's just a previous state by time interaction in the Markov process. The correlations are on a raw data scale. So there're no limits on how high the correlation could be within-patient. There are certain latent variable models where there's constraints in the correlations where you can't really get correlations as high as we need.
- And then the multi-state models give you readily interpreted estimands, including a huge variety of derived estimands. I'll show you some examples. This can be easier to interpret than a competing risk analysis, because death is an explicit part of the outcome. If you want death to not be an explicit part of the outcome, you'll have to relax the proportional odds assumption, which I'll get to. You need more work on this method of analysis if you don't want to count early deaths as the worst outcomes. If you have early deaths that are only possible if the patient was in a later stage of disease when they were randomized than you thought they were and it can't be due to treatment,

then there are some cases in which you may not want to count the early deaths as much, and a time-to-recovery analysis would handle that. This method, with the proportional odds assumption in place, would count the early deaths as really bad outcomes. That's one thing that needs to be considered more.

- What are alternatives? There are marginal ordinal models. Jonathan Schildcrout, who's on the call today, has done a tremendous amount of work with that. Lee and Daniels had a paper in 2007. This is a dual modeling approach where you have one model for the mean process and one for the covariant structure. Having dual models makes it more complex and requires customized programming. That takes a lot of computation time, but the results can be easier to interpret because it is a marginal model—marginal means not dependent on earlier states. It's looking at each day of outcome on its own. Then there's the easiest to implement: mixed effects models, especially in Bayes, that's so easy to do. It's fairly easy to interpret the results until you need to marginalize out certain patient effects. Because it's not conditional on previous states, it cannot handle absorbing or nearly absorbing states well at all, and so that would create an unrealistic correlation structure that would absolutely destroy the operating characteristics of the mixed effects model. There's a link here that goes into a lot more details about model alternatives. There's some input from Lee and Daniels, Schildcrout, and others.
- So let's get to our real task at hand, which is the Markov proportional odds model. We're restricting attention to first-order models. We're going to depend on previous state and not the previous-to-the-previous state. Then the work that I was involved in a long time ago with Bercidas Peterson, she developed a partial proportional odds model to relax the proportional odds assumption. The reason this is important is that you're not going all the way to multinomial logistic regression, which would have a huge number of parameters, which is effectively the way multi-state models are dealt with. They allow all kinds of transition parameters. We're able to relax the proportional odds assumption selectively—for example, with regard to treatment but not with regard to baseline covariates. There's some good software for helping with all this. The very general frequentist VGAM package allows for just about every possible constraint for mixtures of ordinal and multinomial type of things. So it will handle the constrained and unconstrained partial proportional odds model. Then the RMSB package. It will handle the unconstrained model for some purposes, but the constrained proportional odds model for all purposes, and in addition, it handles interval censoring for a given follow-up day. When you're using proportional odds assumption for all the variables, then we're talking about only standard software. For example, if you have a continuous outcome, you could use this method with more than 6,000 distinct *y*-values because the RMS package or M function is made for that. The beauty of the Markov way of thinking is the previous state is just a covariate in terms of how you program it, so you can handle lots of different situations with no special software. Once you start needing the partial proportional odds model, that's when you need something like these. There may be a computational problem if you had 6,000 levels of *y*, in other words, you had a truly continuous variable.
- Now, what I really like about this approach is it handles a large number of special cases. If you had a single time, this is just the ordinary proportional odds or partial proportionalized model. If *y* was binary and represented an absorbing state, then Efron had an amazing paper comparing logistic regression with Kaplan-Meier estimates. The transition odds ratio is almost identical to the Cox hazard ratio. Because in this case, we're talking about a discrete hazard rate. The transition

probability is exactly equal to the hazard rate. If you have daily events that are with a longitudinal set up, the events on a given day have a very low probability, which means that a risk ratio on a given day is almost exactly equal to the odds ratio. Which means they're both equal to the hazard ratio. So there's a strong tie in to the Cox model. If *y* is binary, but there's no absorbing state, that would give you recurrent events. The sum of the state occupancy probabilities that we'll be talking about gives you the mean time-in-state, and the cumulative sum of the state occupancy probabilities gives you the cumulative mean function from a recurrent event analysis. So once you get into this longitudinal way of thinking, you don't really need so many special cases. I've always had a challenge throughout my career, staying up on all the branches of survival analysis, like recurrent events, but most people don't allow for an absorbing state. But then there's newer methods that have recurrent event models with absorbing states, but they're a good deal more complicated than what we're talking about here.

- So here's an example of the quality of the fit you can get from this, using data from that VIOLET study that was so well done. Of course, vitamin D didn't do any good. That's another matter. These are data simulated from the fitted proportional odds Markov longitudinal model. These are state occupancy probabilities as a function of time where time goes from Day 1 to Day 28. I think in this case it's Day 27. If you simulate data off of the fitted model, you get the proportions of being in the different states as a function of time. If you fit the model, the model will estimate the state occupancy probabilities. I think this is one of the better ways to present the model to the clinicians when you're done. They're getting pretty used to these sort of charts and they're really liking them.
- These are the VIOLET-2 results. You can compare that with these above and you see really excellent agreement. The agreement is not just first-order agreement like here, but a much more stringent second-order agreement. This is almost a semi-variogram, where you take all possible pairs of outcomes within a patient and you record how many days apart those two are measured. You can look at a correlation coefficient between two responses within the same patient. You look at that over all the patients. What you have here is the time gap where correlations go down as the time gap grows. That's your typical serial correlation model—AR1 does that—but you see it's not on a single line. The correlations depend on what day you're currently in, not just the gap you're looking at. It's not an isotropic correlation of better like an AR1 model would assume. It's more general than that. If you look at the correlations between the raw data and the simulated observations from our fitted model, you see pretty remarkable agreement. Everything is shifted up. So that's a little bit too high. But overall this pattern is remarkably the same. That's a second-order check.

$$\Pr(Y(t_i) \geq y | X, Y(t_{i-1})) = \text{expit}(\alpha_y + X\beta + g(Y(t_{i-1}), t_i, t_i - t_{i-1}))$$

- So what is the model, after all this buildup? So it's a first-order. Then the measurement times are t_1 to t_m. We're going to just say that this is the MIT, the measurement-at-time *t* for a given subject. So the cumulative probability model for the proportional odds model is the probability of being in state *y* or worse at time t_l. So the probability that capital *Y* is greater than or equal to little *y* for some little *y*, given the baseline covariates and given the previous state, it's just one over one plus *e* to the minus. Now you have a different intercept for each *y*. So that's why this is a semi-parametric model. If this were not expit, if this were *e* to the minus, *e* to the minus, that would be the Cox model. This would be like your baseline hazard function.

Friday, September 3rd, 2021

- **Scott Berry:** So you have a different intercept for each *y*. Does that vary by *T*?
- **Frank Harrell:** It can, and it needs to. And that's a great point, Scott. The reason it needs to, is the mix of outcomes changes over time within patient. We'll quickly mention that this needs to be interacted with time to allow what you said, and I'll give more motivation for that in a second. This is just a lazy way to write this out. There's a whole lot of linear combinations here with other beta parameters. This function *g* is where you spend a lot of your time specifying the model. It's some function of the previous state, the current absolute time, and the gap time between the previous measurement. This would involve possibly a lot of regression coefficients. There is a design headache here—you can't have a design with unequal measurement times and where the gap time is co-linear with absolute time. You'll get a non-identifiability condition if you need to model the gap effect. So that gets into a lot of details I'm not going to get into now, but here's examples of how the previous state could be modeled. You can model it as linear in the previous state. So you just put in numeric codes. You could have a binary indicator for a specific state that has the lowest or highest. You might have a 0 to 100 scale where 100 is a heart attack. You might have a departure once a patient has a heart attack in the previous time. That's special, otherwise linearity in the previous state in other situations. Or you might have a discontinuous bilinear relationship. You have in-hospital outcomes and then you have outpatient outcome severities. So you might have a break in the relationship here with an intercept jump at the transition from inpatient to outpatient. There's a lot of ways to model all kinds of effects that you might be interested in. But these are the most important effects to include in general: the previous state is going to be an absolute must because we do have correlations with inpatients. You almost always need a flexible function of time since randomization. If you have no time effect, that means you're assuming a constant hazard, effectively, like in a survival analysis and we don't like to assume constant hazards. You can have non-proportional odds effects for time. So the mix of events can evolve. You might have patients early on who might get put on a ventilator, but the deaths might occur late. Because the mix of events can change over time, this effect can be very strong so that it would be a need for a partial proportional odds model to allow the evolution of events to re-mix over time.
- Time-by-previous-state interaction, I mentioned this briefly. Hospitalized patients may become more stable over time, so you have an increasing effect of previous state. That's easy to handle. You might have a flexible function of gap times if you have unequal measurements. You may have an interaction between previous state and the gap time. You might have an interaction between time and treatment if the treatment effect is delayed. This would be your classic time-by-treatment interaction consideration that any model would have, whether it's time-to-recovery or what have you. Now in terms of treatments, since treatment is special, if you're assuming proportional odds—so the treatment has the same effect over all levels of outcome—then treatment will have a single coefficient in the transition model. Remember the probability we're calculating is the probability of a transition or the probability of being in a certain state given your previous state. There's a single treatment parameter for all purposes, whether you're talking about derived estimands or the primary odds ratio, and information is being borrowed across all levels of *Y*.
- Now if you relax that assumption, you can use the partial proportional odds or constrained PPO model. If you're using a frequentist approach to doing this, there's no possibility that I know for borrowing of information across the relaxed categories. That will cause a pretty big power loss. In other words, to have a mortality statement, you need to power this study for mortality as if that's

the only outcome. But with the Bayesian approach, you have the ability to put a skeptical prior on the departure from proportional odds. How skeptical that prior is tells you how much information you're going to allow to borrow. Then you have this really nice ability with Bayesian models to compute the probability that the treatment effect on mortality is more than epsilon different than the treatment effect on nonfatal outcomes. Just think of all the studies you've ever been involved in where a clinician was interested to know about consistency of effect. Well, here you have a very formal way to do that. Most of the methods that have been used before are pretty informal.

- Now, what we really need for interpretation is to go from transition probabilities to state occupancy probabilities. If you have equal time spacing the notation is easier. So the probability of being at level *y* given the baseline and only the baseline—not given the previous state—you have to uncondition on the previous state. You just have this log total probability and it'll go back one time point. Then of course, this depends on T minus 2 and so on. It's just a recursive matrix multiplication. The really beautiful thing about this is this calculation is an after the fit calculation. You'd never do this while you're doing a fit. That's really important in a Bayesian setting where you might need 10,000 posterior draws to get all the chain convergence and everything, or say 4,000. These are calculations that are done on demand after the fitting is totally finished.
- So you would do this for the covariate headings you're interested in or for every one and it wouldn't take that much time to do it. That will give you a semi-parametric unconditional distribution at *y* at each time. This is more like what people are used to thinking about. One of the points I want to make here, which I think will be considered as radical to some people, is a lot of people like to use mixed effects models, because each time point it's easy to interpret, but they're actually quite restrictive. A lot of people avoid state transition models, because they don't instantly give you these unconditional probabilities; it's conditional only on the baseline. You can get around that so easily. Then your unconditional distribution is very flexible. It could end up being a bimodal distribution or something very non-normal. We have software functions in R that make these calculations very easy to do for frequentist and Bayesian. And in the Bayesian world you have a very special advantage: if you were trying to use the Delta method to get a confidence interval on a state occupancy probability in the frequentist world, you would go to a heck of a lot of trouble because it's such a complex nonlinear transformation. Then you'd be disappointed in the confidence covers because you find the confidence intervals need to be asymmetric. The Delta method doesn't allow for that.
- Contrast that with the Bayesian approach, you're going to get your posterior draws on the primary parameters, which is the betas and alphas. Then for each of your posterior draws, you're going to calculate whatever derived estimand you want. You'll compute, say, 4,000 to 10,000 values of that. If one of the things you derived is a state occupancy probability on time 3 or mean time, in state, you would just calculate these quantities 4,000 times and get your highest posterior density interval and your posterior probabilities and you're done. You're not using any approximations other than your Monte Carlo error. And you don't have to learn the Delta method. You're not going to force the intervals to be symmetric at all. So they'll be accurate.
- So what are some of the estimands? This is your primary estimand: transition odds ratios. You can get transition probabilities that are conditional prior state or covariates. You can get covariate-specific state occupancy probabilities, time in a certain state such as state 3, time in state 3 or worse. That might be interpreted as the mean time unwell. For example, if *y* less than 3 was well,

you can get differences in mean time in states and get a Bayesian intervals for those easily and all kinds of others. So whatever you name you can get it. We have a graduate student, Max Rohde, who's doing a lot of nice work on this to show that when you have the proportional odds assumption for treatment (and he's applying this to ACTT-1 data for remdesivir), you have tremendous agreement among posterior probabilities across all the estimands when the thing you're assessing is whether there's any efficacy or not. If you just want to see if treatment is moving patients in the right direction, if you have a proportional odds assumption, it doesn't matter what estimator can use for that.

- What gets interesting is when ε is greater than 0. You want to see what evidence we had that there's at least a one-day improvement in the time-in-state. And then what's the inference for at least improvement by two? So once you get into levels of efficacy, it does have a different interpretation for every estimand, and even for estimands on the same scale, you'll get different absolute efficacies for high-risk covariates versus low-risk covariates. I think it's a good way to look at what is your decision estimand and does it matter. And for most purposes, it will only matter if the proportional odds assumption is badly violated and you allowed it to be violated.
- So this gets into assessment of fit. For the right-hand side of the model, we have all of our traditional assumptions that we relax using regression splines and adding interactions. It's just that some people are not used to thinking of time so prominently in a model, but you can have interactions with time, which is, for example, how you relax the proportional hazards assumption in a Cox model. You have the assumption of proportional odds for treatment. You can look at variation of cutoffs, specific treatment odds ratios. The trick is when you start exploring the observed patterns, we don't really know how to include our subjective impression of that exploration in our uncertainty assessment. There's something to be said for a more cohesive approach where you agree on a prior for the amount of non-proportional odds. And that will dictate how much borrowing you have and how much sample size you need. You just stick to that. You don't actually look to see if proportional odds was violated. You don't try to say that it wasn't violated. But you have something like this that would respect the fact that when your sample size is 40, the prior effectively is going to dominate. You're going to be borrowing information and if the sample size is 400, you're not going to be borrowing much information across levels of the outcome. For the correlation structure, we talked about variograms. This is another thing that's really easy to do. When I did this for the VIOLET study, I took the Markov first-order model, proportional odds, and added a patient-level random effect. I was able to show the random effect variance went down by about a factor of 10. If you fit a random effect model on the original data, not using the transition model, you get this massive random effect variance, which is really in this case an indication of a problem with the random effects model. But when you assume conditional independence, because you're modelling transitions, the variance of the random effects is very close to 0. This I do as a test of the conditional independence assumption in the Markov model. It's very easy to do. Then the checking of dependence on the previous state, that just really goes into the right-hand side of the model again because there's nothing really special about the previous state. Then the most general way to do this, which I need to get more experienced with, is to do posterior predictive draws, which, before I got more Bayesian, I just called that re-simulation. Check your simulated data against the raw data. A lot of resources here that you can go to.

Friday, September 3rd, 2021

- So I think I'm going to stop there and see what kind of discussion and questions people have. Maybe we can start with Nancy, great question [in the chat]. "The downside of the great flexibility, possibility or choices are incorrect." I think that's always a great point and it's always happening, Nancy. I'll give you one example. If you look at ACTT-1, the treatment is not in proportional hazards, but the published analysis for ACTT-1 assumes proportional hazards and it gives you a very simple result in a single p-value. But how do you know what the effect of the non-proportional hazards is, when you add flexibility, which you'll see in our ACTT-1 re-analysis? When you add flexibility by allowing interactions with time, how are you going to specify those? You have great flexibility. Are you going to do a linear interaction which will not lose much power if you have proportional hazards? Are you going to do a spline interaction to allow for flexibility? And what I show in that analysis is that when you allow non-proportional hazards and you allow it to be very flexible, you actually lose power. So everything can happen.
    - **Nancy Geller:** Yeah, my concern is, how many times are you going to do this? I mean, you make the best assumptions that you know, and you do it. Then you do this assessment of fit and it fits okay... But what if some of your assumptions could be improved upon? The problem with complex models is that you have so many choices. How many times do you want to do this to make sure you get it right?
    - **Frank Harrell:** I think I have a good answer to that and a weak answer to that. I think the weak answer is that if somebody thinks your model is too simple and they convince you and the people who are going to review your protocol revisions to generalize that model, that will be self-penalizing, your posterior distribution will get wider, your standard errors will get wider. So you can work in a certain direction and be okay. But to get to the stronger answer to your question, if you look at our links here, we have reanalyzed VIOLET, we have reanalyzed ORCHID, which was our PETAL network hydroxychloroquine study, we've reanalyzed ACTT-1. I think all three of these studies serve as models for our studies that are currently in the design phase—with the exception that we're featuring more outpatient data that we're less familiar with. My strong answer to your question is, when you have the luxury of great datasets that other people have collected with completed clinical trials who were willing to share the data, this is your very best opportunity to learn about the dominant patterns that you need to not forget because then you have complete pre-specification. That's where I learned the hard way that if you assume proportional odds for time, you can destroy the inference. If the mix of events changes over time, and you assume it doesn't, you can get huge alpha inflation in this model. Huge.
- **Scott Berry:** So first Frank, I love this, I think it's tremendous. So thank you for that. A couple of comments and a question. I think what becomes the real challenge scientifically here is, the really important assumption about the medical aspects of this is that you have ordered states—this state is better than that state. And how granular you get when you model that. If you model things that aren't medically important, but they're ordered, that can drive the outcomes. But I think that's actually where statisticians need to be. We need to be working with the clinicians to get that right. I think this general framework allows us a really, really nice way to do that. I think others are going to find that undesirable, but I actually think it's where we need to live. Now a question I have for you: an assumption in this model is that you measure the status of a patient on day t without error (they're on organ support, they're in the hospital but off organ support). We have a number of scenarios where you want to do state-space models like this, but you get an outcome that's

measured with error (what is their cognitive status on that day, what is their Modified Rankin, but yet you still want to use the underlying state-space aspect. Have you done any thing where you do state-space modeling, but the measurement itself has error?

- **Frank Harrell:** I have not. The only thing I've done is in a univariate case is just, I have a little R-studio simulator that allows you to add error and see what happens. What it shows is there's no amount of error that would make a dichotomous analysis work as well as an ordinal analysis or a continuous analysis. There's various simulations you could do, and we need to generalize that to have a better simulation model. But I view measurement error here as very similar to the impact it would have in mixed effects models and all the others—which is to say it does have an impact, it reduces our power, and sometimes there's not anything we can do about it. Always a good point to worry about though. There is one complication I haven't figured out yet: if you're making this conditional probability and in the previous state, it was interval censored. This $Y(t_{i-1})$ was interval censored in the previous day. I don't really know how to condition on a partial conditioning there. I've just taken a midpoint as the conditioning event.
- **Scott Berry:** What if you did latent variable multiple imputation for the unknown status of the patient?
- **Frank Harrell:** Yeah. You could do regular imputation and posterior stacking or something, but I'm just trying my best to avoid that for the moment. The likelihood function is enough to work with.

- **Daniel Rubin:** Thanks for the great work. My question is a follow-up to Nancy's, and that's just for this function *g* in the model of the previous state and the time, I was wondering if the software has defaults that you would recommend for common types of outcomes or data structures? Or I guess what I'm trying to get at is how much hand crafting is required to use this method, for an individual trial, or to what extent can the method be used off the shelf?
  - **Frank Harrell:** First of all, let me say that most people that are doing longitudinal analysis are using stuff off the shelf and not even realizing the severe assumptions that the off-the-shelf method makes, like SAS PROC makes. It really doesn't fit most longitudinal datasets the way people are using it. But what we're talking about here takes more thinking than the univariate case. But the thinking is really that you have a more glorious set of baseline covariates. These are all messed with the same way you would mess with baseline covariates. I think this is my checklist here of the things you need to figure out to specify this thing up here. These are what I've found to be the dominant effects that you don't want to forget. I think when you get into your own study that has different considerations and different outcome properties, you'll refine this, but I would encourage you to just think about this as your initial checklist in specifying. Then that's going to all go into your covariates. The software just will see this as regular covariates. The only thing the software sees as different is if you have 1,000 patients and on the average, they have 10 days of data each, you're going to have 10,000 records in your dataset.
- **James Troendle:** I had a similar concern. I think it's great—really my only concern is really what Dan was asking about with the *G*. You say it's like adding baseline covariates, but do you have the same kind of protections of Type I error though if you don't get the *G* right, as you might in a Cox model

where you might have attenuation if you don't get covariates in there that are important? Do you have any kind of protection like that or for Type I error?

  - **Frank Harrell:** Well, I think it's a broader question than Type I error. I would worry about Bayesian power, for example, if the model is misspecified, but it could affect β error or α error, just like any model can in my experience. One of the handouts that's very long shows this; where the Type I probability can be messed up the most is when you have the most prevalent outcome changing at different follow-up times. When you have more than two outcomes, so that would be the non-proportional odds for time. Other departures are not going to mess up the operating characteristics as much as that. Unless you had a very severe non-proportional odds for treatment that you didn't model at all. But I think this model is really not different from others. Just like if you have no proportional hazards and you don't have any time-by-treatment interaction in the model, some stuff's going to happen. I don't see this model is magically different from a lot of our models that we use in the regard that you're asking about. I forget to stress the previous state. If you don't model that correctly that can have a pretty massive effect. That is another thing to pay close attention to. If you model the previous state as a linear effect, that's very restrictive. If you model it as categorical—in other words, you saturate the model with regard to previous state—then you're going to fit previous state really well. And the question is, do you need an interaction between absolute time and previous state?

- **James Troendle:** Yeah, I was wondering about that time. You went very fast over Scott's original question and how it had to do with the interacting alpha *Y* with time. But it seems complex. I don't know that I really understood that so fast.

  - **Frank Harrell:** This notation hides that, but α_y, little *y* only comes into play in this. So you'd have one intercept for each distinct *y*, less the first one. But what's hidden in that is if you had your treatment variable here and let's say at β_1 for treatment, then you might have a separate β_1 for whether *y* is greater than or equal to 2 versus *y* greater than equal to three. So that's the partial proportional odds model. It's the same way for time. Instead of having a simple time effect here, you can have the time effect has a different impact on *y* greater than or equal to two, than it has a *y* greater than or equal to three. If you go to the links, I've written that out in more glory. Where you actually see the non-proportional odds parameters, which I think I call τ in some of the other handouts.

  - **James Troendle:** That seems complex.

  - **Frank Harrell:** When we see something is complex, always think about how the way we're analyzing data now looks to be more simple but is only hiding complex things.

- **Scott Berry:** Frank, I agree with you on that point. It seems like in a lot of the endpoints you're looking at that if you're doing daily values, the likelihood of being in the same state you were yesterday is high. That that α_y could almost be a mixture where there's some likelihood you just don't move, would fit that really well. But that's also a transition that you have a ton of data on. And you can be very flexible in modelling that, where the treatment effect is a little bit different. I think you have a good bit of power to estimate the transitions as a function of last state. The one thing you seem more willing to do then I is allow non-proportional effects. Because I think you open up the huge issue here—which is even more complicated—of what states are. You almost get stuck

in a utility analysis of states. If I have differential effect across this scale, I may not know which treatment is better.

- **Frank Harrell:** Yeah. And that's a little bit more with regard to non-proportional odds on the treatment effect than it is on the state effect. But for your earlier comment about the previous state effect, I'll do everything I can to try for proportional odds for the previous state effect. Because otherwise the model gets to have a whole lot of parameters. But I'll model the main effect of previous state in a saturated way. And then the question is, are you going to relax the proportionality assumption for treatment, and I'll work hard not to do that. In some cases, if the clinicians who have a say in this feel really strong about it, but not super-strong, the best approach is going to be to have proportional odds for treatment borrowing maximum information with a highly featured secondary analysis where you relax that and give additional clinical readouts to the outcome-specific treatment effects.
- **Scott Berry:** I agree with that.

- **Eric Leifer:** You mentioned you did an analysis of ACTT-1 data, but I haven't yet read up on what you wrote there. I guess in ACTT-1, the primary outcome was time to getting out of the hospital, if I recall. I was curious what direction your reanalysis goes there, in terms of adding different states.
  - **Frank Harrell:** Yeah, we might want to schedule that for like a 15-minute presentation at a future meeting, because I think it's a very interesting question. The short answer is, the proportional hazards assumption is violated and the treatment effect of remdesivir takes a few days to have an effect and then it wears off after about Day 12. If you use this transition model, you'll see that more easily. You'll be able to model the treatment effect duration a little more visually or interpret it easier. You'll see that the treatment has maximum effect at Day 10 and then it starts to wane. So this really has a lot of ramifications in terms of what should be the summary statistic.
  - **Eric Leifer:** When you say the treatment is more effective at Day 10, what does that mean in terms of time to getting out of the hospital?
  - **Frank Harrell:** They're getting a kind of average hazard ratio, which I think gives you the right answer in terms of directionality. It doesn't necessarily give you the right answer in terms of magnitude of efficacy. In the transition model, there is a time-by-treatment interaction. That is that it's just another way to look at a non-proportional hazards.
  - **Eric Leifer:** I'm having difficulty going from one to the other, sorry.
  - **Frank Harrell:** It's really weird, so don't apologize. I encourage everybody to look at the resources. If you have any follow-up questions, just email me. I'm hoping we can come back to the ACTT-1 study re-analysis at some later time, at least for a brief discussion.

Friday, September 3rd, 2021

Chatbox

Stacey Adam 09:03

HI All, We will start soon.

Stacey Adam 09:15

Just waiting on others to see if we get more joining

Stacey Adam 10:53

Lisa is unable to join today and Karen will be a little late, but we will get started with Frank's

presentation shortly

Nancy Geller 35:51

This is great work, Frank! The downside of the great flexibility is the possibility that your choices are incorrect. Comment?

Chris Lindsell 52:12

Agreed Scott - completely!!!!

Kevin Anstrom (he / him / his) 01:06:38

This is great work. Thank you Frank