Jill Harrison, PhD:

Hi, this is Jill Harrison, executive director of the National Institute on Aging IMPACT Collaboratory at Brown University. Welcome to the IMPACT Collaboratory Grand Rounds podcast. We're here to give you some extra time with our speakers and ask them the interesting questions that you want to hear most. If you haven't already, we hope you'll watch the full Grand Rounds webinar recording to learn more. All of the companion Grand Rounds content can be found at impactcollaboratory.org. Thanks for joining.

Hello everyone. I'm Jill Harrison, one of the executive directors at the NIA IMPACT Collaboratory. Welcome to today's podcast. I'm joined today by Dr. Michael Harhay. Dr. Harhay leads a research program that develops statistical methods, study designs, and outcome measures to improve studies for patients with critical and severe illnesses. As director of the NIH and PCORI funded Clinical Trials Methods and Outcomes lab at the University of Pennsylvania's Palliative and Advanced Illness Research Center, his team is particularly focused on developing and using causal inference and Bayesian statistical methods to better design and analyze pragmatic and cluster randomized trials for these patient populations. He's contributed to a wide range of trial methodology areas including informatively missing or truncated data, Bayesian trial applications, small sample corrections, and making the best use of baseline covariate information.

In 2022, he joined the Penn Medicine Nudge Unit as the director of statistical evaluation, where he applies these methods to studies in Penn's health system. Dr. Harhay currently serves as deputy editor of the American Journal of Respiratory and Critical Care Medicine, and an associate editor of the International Journal of Epidemiology. In 2023, he was named an honorary research fellow in the MRC Clinical Trials Unit at University College London. Dr. Harhay, thanks so much for joining me today.

Michael Harhay, PhD, MPH:

Yeah, thank you for having me.

Jill Harrison, PhD:

Of course. This is the companion podcast that accompanies a presentation you made last week at our monthly Grand Rounds on the topic of outcomes truncated by death in randomized trials. Can you give our listeners tuning in today a high level summary of your talk and how it relates to clinical trials of non-drug interventions in dementia care?

Michael Harhay, PhD, MPH:

Yes, of course. An unfortunate reality is that individuals who are randomized into trials do not always complete the trial. This can occur for a wide range of reasons, such as no longer wanting to be in the trial, moving away, or an unfortunate event that I discussed during my talk, they die. When they die, the balance that we achieve when we randomize interventions to two different groups of people can start to be eroded and some of the decisions that are often made based on what statistical model you use, how you deal with that missing data, can conflate and work together to complicate the treatment effect estimate that you get out at the end of your trial.

The goal of that talk that I gave was outlining various considerations and solutions that are available to researchers, none of which are necessarily optimal, but try to work through the different trade-offs that each one offers us so that there's a little bit of a framework that researchers could work through in their own trials, or just when they're reading the literature to try to understand what they're reading.

Jill Harrison, PhD:

Excellent. Well, walk us through some of those trade-offs, please, as we go through some of the questions that we didn't get to during the Grand Rounds broadcast. We have a question that was posed in the chat. Can you briefly describe how these analyses might affect generalizability of the study results?

Michael Harhay, PhD, MPH:

Yeah, so it's a great question. The idea of generalizability is that what you learn in one trial, or in the setting in which a trial is executed, is applicable in another setting. The challenge that occurs is that some of these decisions, for instance, making a composite outcome, a common composite outcome is that you will assign people a quality of life score if they died of the worst value. If you have something like the EQ-5D and it is on a rank of one to five, everybody who dies will get the lowest value or a negative value that represents some stage of worse than being alive. That is helpful from a sense to get the net benefit and net interpretation of an intervention in that specific trial. But, what you're doing there is you're fundamentally changing the value that that outcome is taking on, and as a result, it doesn't functionally mean anything, or it doesn't mean that much.

It's difficult to think about what is a quality of life scale of negative one. If you look at the mean value of quality of life, it will start to shift that mean value and it could make a mean value that's not really representative. It is, in a way, a solution that allows us to estimate which arm in a clinical trial does better than the other arm, but the translation and what that means to an individual, so you can't just say, "If you take this intervention you're going to see an average increase of quality of life of two points," because that actual value of two points is kind of complicated. It's this mixture of a ranking we have essentially applied to the data to solve one issue. It's kind of always a give and take with a lot of these issues, is you solve one problem but at a trade-off of another one. Generally what you're solving is a missing data problem. You are creating a solution to fill in the data you have, but often at a conceptual trade-off.

Jill Harrison, PhD:

Speaking of that EQ-5D measure, we have a question on that specifically. Can you comment on strategies or approaches on the handling longitudinally collected patient reported quality of life outcomes in the presence of death or other intercurrent events such as when the EQ-5D score of one is often viewed as full health and a score of zero is interpreted as a state as bad as death?

Michael Harhay, PhD, MPH:

That is a very common approach and it is this desire to create what is, in a sense, a ranking. Everybody who's alive has some value that's better than someone who died, and by applying a value to those who died, they are now no longer having missing data. The cherry trade-off is the one I just walked through, is that now we have a scale that was made to rank from one to five, but now it's ranking from zero to five. Just thinking about what the mean, median, even the minimum is, and then if you put that into a regression model and you try to compare arms, it just becomes complicated. Nevertheless, this is a very popular approach that's used throughout many subspecialties of medicine as a solution, and I don't necessarily think it's the worst approach.

Of the different approaches out there, I think it's perhaps the most attractive. To give you just kind of a trade-off, an alternative approach that one could use that is very popular, people use this thing called inverse probability weighting, which is a technical term that generally means that you downweight the outcomes that were observed among people who died. You try to make them matter less in your

treatment effect estimate, and you upweight the people who survived. But, what you get there is an estimate, and the estimate is interpreted as a hypothetical estimate. It's what we think we would've seen if no one had died, which is kind of a weird concept. Even though the application of a value of zero doesn't necessarily make sense, it does get around, I think, the problem and it still gives you a little bit more of a sense, at least of a distribution of health, and kind of acknowledging that death is the absence of health.

Jill Harrison, PhD:

Very interesting. This is a question we have from Thomas Bayer. "Are there any important problems with hierarchizing death with hospitalization and other burdensome forms of care versus the DAOH composite that you presented in the Grand Rounds broadcast?"

Michael Harhay, PhD, MPH:

Hierarchical endpoints, I think, are attractive because they let you weight outcomes as more or less attractive in a scale. I think they are tricky to always figure out if those are shared values. There is this belief that people want to be outside of a hospital, which makes, I think, a lot of attractive sense, but maybe someone benefits from an institutional setting. Maybe they don't feel as lonely. Even though this may be a rare instance, those valuations that we apply in how we say this is a better outcome than this outcome may not be universally shared. This is where I think the science of hierarchical endpoints has a lot of ability to be very tailored to specific patient populations, but that really involves what is, to date, I don't think a heavy stakeholder input. A lot of these are tended to be ranked and valued based on clinical investigators. Not to say that they're necessarily bad, but bringing a little bit more of a patient presentation or even just caregiver perspective into these, I think, can help us do better.

But at the end of the day, once you say this is better than that, there's probably someone in the world who doesn't fully agree to that. Trying to figure out how to balance those set of trade-offs is, I think, the biggest challenge. I think the bigger challenge that people get is they get excited about hierarchical endpoints because they think they can put multiple outcomes into them. That is true, but by adding two to three to four outcomes, now you're trying to think about all these different permutations, so that ranking becomes a little bit more built into how you interpret an intervention as effective or not, and it's easy to criticize them on the other side.

I see a lot of trials that people say "That's not how I would've constructed it. I think that's too complex. I wouldn't have ranked this the second outcome, it would've been third." It's kind of the ongoing tension I mentioned at the beginning. These are all solutions that are kind of not optimal. I think hierarchical outcomes have the ability to be the most flexible and most tailored, but making them kind of optimal still, there's a way to go and a lot of research opportunity there.

Jill Harrison, PhD:

Has there been any type of partner engagement or people who are experiencing the disease that is being studied? Have they looked at any of these hierarchical outcomes and rank ordered them or given feedback about their preferences or their interpretation of their usefulness and value?

Michael Harhay, PhD, MPH:

To my knowledge, this is pretty limited, I think in a sense, because the hierarchical outcomes themselves are relatively new to the scene. Not yesterday, but in the last five to ten years, they were largely advocated. The first time I saw them was in the cardiology setting and now I feel like they're advancing. In the palliative care, in the older population, we have a colleague at PAIR, Dr. Katie Auriemma, who's

been working with patients longitudinally with respiratory failure and also palliative care and trying to understand if they can, basically, quality weight days.

This idea of looking at having people fill out EQ-5D, frequently longitudinally, and trying to see if we can get a little bit more granular and a little bit more insight than just saying out of the hospital or not. I feel like that is emerging science and figuring out how to capture those quality of lifes in a pragmatic way, working with phone surveys and making it easy to capture but also informative has an enormous amount of opportunity. If I was a researcher getting into the field, I can't think of a more patient-centered and clinically relevant and informative endeavor. I think there's a lot of opportunity there for young researchers.

Jill Harrison, PhD:

Well, speaking of trialists and researchers that are interested in this topic and in the field particularly of dementia research, which is the primary focus of course of the NIA IMPACT Collaboratory, what training resources would you recommend? Where should people start? What article should they read, books, et cetera, webinars, you name it? How could we open the floodgates for them?

Michael Harhay, PhD, MPH:

There's a couple of readings that I'll just mention within the sense of trying not to overwhelm. There is a BMJ, British Medical Journal, article from 2018. First author is Elizabeth Colantuoni, and the title of the article is "Statistical Methods to Compare Functional Outcomes in Randomized Controlled Trials with High Mortality." Though not specific to dementia, I think that this is a general challenge for a lot of different disease and patient populations, and that's a very nice article that goes through many of the considerations I mentioned the other day.

We also have an article that I'll happily plug. It is a 2023 article, and it's in the American Journal of Epidemiology, and this is a little bit more technical. The one I mentioned before has a statistical focus, but is more general oriented. If you're interested a little bit more in a deeper dive into statistical trade-offs that I brought up during the talk, our article, the first author is Brennan Kahan, K-A-H-A-N. It's "Eliminating Ambiguous Treatment Effects Using Estimands," and it's one of the articles I showcased in my article, that's in American Journal of Epi.

If anybody wants to get much deeper, I'm always happy to field requests. I have a deeper dive, and the last one I'll just mention is a more recent article in BMJ Open, and it's called "Estimands: Bringing Clarity and Focus to Research Questions in Clinical Trials." I think that's a nice survey from conceptual to statistical complexity and detail that should give a good framework. There's a scatter of literature, I think, from historically different approaches to this, and I feel like those articles are starting to synthesize it and try to create more of a generalized framework to think through this issue more broadly.

Jill Harrison, PhD:

Excellent. Thank you so much for making your offering to be available to listeners that are interested in this topic. They can find Dr. Michael Harhay at University of Pennsylvania. Dr. Harhay, thank you so much for sharing your work with us and your time today.

Michael Harhay, PhD, MPH:

Thank you, my pleasure.

Jill Harrison, PhD:

Thank you for listening to today's IMPACT Collaboratory Grand Rounds podcast. Please be on the lookout for our next Grand Rounds and podcast next month.