Jill Harrison: 00:02 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@impactcollaboratory.org thanks for joining.

Vince Mor: 00:30 This is Vince Moore. I'm one of the multi PIs along with Susan Mitchell of the new NIA funded impact collaboratory effort designed to improve the quality of life of persons living with dementia and their caregivers by introducing interventions that we think might work because they'd been shown to do so when researchers implement them and to see if they can actually be implemented in a functioning healthcare system per se. And today I'm delighted to introduce Monica Taljaard, who is a professor at the University of Ottawa and Ottawa Hospital, and a specialist in statistics of cluster randomized trials. And we're asking her some questions following her great webinar, focusing on the step wedge design, which is one kind of a cluster randomized trial.

Vince Mor: 01:24 So Monica, that was a wonderful talk. I just wanted to ask a few questions here. So you mentioned during your talk that to properly calculate the sample size needed for step wedge design, the investigator needs to estimate the correlation of the outcome over time. Is that the correlation between successive measures of the outcome variable at the person level or at the cluster level? Could you help me that?

Monica Taljaard: 01:53 So that is a great question, Vince. Sample size calculation procedures for the step rich design are more complex precisely because we need to account for these more complex correlation structures. So not only do we need to account for the regular intercluster correlation coefficient or the ICC that we're all familiar with, but we also need to account for an estimate of this correlation over time. Exactly as you mentioned now. So your question is whether this correlation or the time is at the person level or at the cluster level. And the short answer is it can be both depending on the type of stepped wedge design. So if we first considered the cross-sectional stepped wedged design, so that is the design where we have measurements taken on different rather than the same individuals in each period. And suppose we've got a continuous outcome. So if we're planning to use a mixed effects regression

analysis approach, fortunately we have the sample size calculation available and they do indeed require us to specify these two types of correlation coefficients.

Monica Taljaard: 03:08 So the first is called the Within Period ICC and this is simply the correlation between multiple individuals measured in the same cluster and in the same period .it's the regular ICC we are all familiar with. Although it is attached to a specific length of time, don't forget that. So the length of a single step in a step which design.

Monica Taljaard: 03:30 The second correlation coefficient, which is new, is called the cluster auto correlation coefficient or CAC. And this is the one that can be interpreted as the correlation between repeated cluster means over time. But I just want to mention that there is an interesting alternative way of thinking about this CAC or cluster order correlation coefficient and that is that one minus the CAC represents the percentage decay in the strength of this within period ICC over time. So for example, a CAC of 0.6 represents a 40% decay in the strength of the ICC when the individuals are observed in different steps rather than the same step. So for example, if they're within period, ICC is .05 and we assume a CAC of 0.6, it assumes the between period ICC is 0.03. So it makes perfect sense that the between period ICC is less than the within period ICC because in most longitudinal studies, the strength of the correlation tends to decay over time.

Monica Taljaard: 04:48 So I've just explained these two types of correlations for a cross-sectional stepped wedge.

Monica Taljaard: 04:55 Now for a cohort stepped wedge that's a design in which we take these repeated measurements on the same individuals over time. In this case, not only do we have this Within Period ICC and the CAC to specify, but we also need to specify a third correlation. That's the person level correlation over time and that correlation we call the Individual Auto Correlation coefficient or I see and we can think about this correlation simply as the strength of the correlation in repeated measures on the same individual. So just in summary, so for both the cross sectional and cohort designs, we need to specify these two ICC parameters. The Within Period ICC and the CIC. And then for the co design we need a third parameter, namely the individual auto correlation coefficient as well.

Vince Mor: 05:57 That's great. Now just a little clarification. So because many investigators are familiar with the correlation of measures over time within the same person as you've just discussed, but less

so at the cluster level. Do you have any guidance for how one gets those cluster level estimates, whether it's for the cross-sectional or the longitudinal, the within person over time?

Monica Taljaard: 06:21 Absolutely. So there are several things we can try to do to obtain estimates for these parameters in advance. And the very best advice I can give is to try to obtain raw historical data for your trial. So ideally for a similar target population, similar clusters and similar time intervals as for your actual design. Now fortunately step rich designs are often done in settings where the outcomes are assessed using routinely collected data. So that's an ideal position to be in because then you can try to gain access to some historical data in advance of the trial and you can feel really confident in your sample size calculation.

Monica Taljaard: 07:04 So assume first we have a continuous outcome. So what we normally do is fit a linear mixed model to these historical data and we will use the same random effects as what we would use for our actual analysis. Of course, we need also a fixed effect for time, but there's no treatment indicator. This is merely observational data. And then we just take the estimated variance components from that mixed effects regression analysis and we put it together to determine the Within Period ICC, the CAC, and in the case of the cohort design, the IAC.

Monica Taljaard: 07:42 Now I just want to say something. I've been talking about continuous outcomes all along. What about binary data? Most of our step wedged trials usually have a binary data, so in that case we recommend that we still use a simple linear mixed model for estimating these parameters rather than a random effects logistic regression model. And the reason is that our sample size methods require these ICC estimates to be on the proportions scale. But if we do a random effects logistic regression model, all of our estimates are going to be on [inaudible 00:08:24] scale and unfortunately, we have no easy way to convert from the Loge at scale to the proportion scale to get the right scale for those ICC estimates.

Monica Taljaard: 08:35 And I should also just say that we don't really know if this method works very well for binary data. We know the linear mixed model works really well for continuous data, but we don't really know how to estimate them definitively for binary data. In the absence of anything better, we recommend just using a linear mixed model, but this is definitely in an area where we need more work.

Monica Taljaard:	09:02	So what do we do if we don't have any routinely collected data available? I would say at the very least you should try to obtain a reliable estimate of the Within Period ICC, so from previously published trial reports. And then perhaps you can just hazard a guess as to the extent of the correlation decay to get an estimate of the CAC. So generally perhaps to be conservative, you can just set your CAC at 0.5 or 0.6.
Monica Taljaard:	09:36	Now if you don't have any routinely collected data available whatsoever, nor do you have any previously published estimates, it's not a good situation to be in. But even here, we've still got some rough rules of thumb we can follow. For example, we already know that ICC is for clinical outcomes tend to be smaller than four process measures. And we also know that ICC is for binary outcomes tend to be smaller when a prevalence is low and when the cluster sizes are very large, the ICCs tend to be smaller. And I would also hazard a guess that ICCs also tend to be smaller with longer time intervals.
Monica Taljaard:	10:18	So these are some rough rules of thumb for the ICCs. We don't have any explicit rules of thumb for the CAC. And then just a final word of advice is please for investigators to publish these estimated correlations from the completed trials because that could help another researcher plan their trial.
Vince Mor:	10:42	That's actually great advice. Just a question of getting the editors to allow one to put it all the information you'd like to have, but that's actually great.
Vince Mor:	10:50	So next question, also related to power, is in regular cluster randomized trials, a conditional on any given ICC value, the number of clusters increases the power more than any increase in the number of subjects within a cluster. Does the same principle seem to apply in step wedged designs and or is there some other hidden problem?
Monica Taljaard:	11:15	Yeah, so with step wedged cluster randomized trials, we still get more bang for the buck, if you will, by increasing the number of clusters rather than the number of individuals within clusters. So it's always preferable to have more clusters and enough, of course, as you know, this is because of the presence of the intercluster correlation. So there's diminishing returns to increasing the cluster sizes. But now there's another way to think about this in a step wedged cluster randomized trial, because we have this design, we often take measurements on all the available within a cluster. So further increasing the number of individuals within a single cluster period is usually

not possible. But it may be possible to improve your power in such cases by increasing the number of periods. So the number of measurement times, which essentially increases the duration of your study.

Monica Taljaard: 12:14 So for example, suppose you've got 10 clusters and your planned step lengths which you've worked out based on logistical considerations are three months. So one possible design is to have two steps and three periods. So that gives you a total study duration of nine months. But you can improve your power here by increasing the number of steps and therefore also the study duration. So ultimately, you will get the most power by leading each cluster cross at its own step. So in this case you may design your trial with 10 steps and 11 periods, but of course, that increases the total study duration to 33 months. So you will have to consider whether that is still logistically feasible. And in general, we find that the power increases most from increasing the number of steps from a minimum of two to around six steps and beyond about six, maybe eight steps, the power benefit from increasing the number of steps further starts to decline. So that's a simple way that you can try to increase power when you've only got a limited number of clusters.

Vince Mor: 13:28 So that then balances to the next question. Of course, your duration of the study is always then increasing your risk of the potential for confounding with time. And what happens when history intervenes in the midst of a step wedge design in a way that you suspect could affect how this intervention is implemented or the measurement of the outcome. And is there some analytic recourse that you would suggest or that you suggest... Yeah, that other people can take?

Monica Taljaard: 14:00 So that's the final question, which I'm hoping we don't have to answer. I think the fourth question will be more useful for us to consider Vince. That question is actually, I don't think there is any analytical recourse because if there is an interference that affects all of the clusters at the same time, in that period, you only have clusters exposed to that interference. There's no control in that period. So I'm pretty sure your analytical model is going to fall apart.

Vince Mor: 14:32 Okay. So then the earlier question was about measured outcomes. So when you're... Often investigators care about the measured outcomes even though they're conducting a pragmatic trial because they want something, for instance, in the case of dementia, about the caregiver, and very rarely is

routinely information collected about the caregiver. So you want to have a measurement to how the caregiver fields, and this has got to be collected in some way by the researchers, but measured outcomes are often subject to missing data. Is this analytic challenge complicated when conducting a step wedge design even more so than a standard phase three per person level randomized trial?

Monica Taljaard: 15:16 I think so. So missing data is a challenging problem of course in all types of trials and no less so in cluster randomized and step wedge trials. So I'm not aware of any work that's been done on the topic of missing data in step wedge trials specifically, but I think we can probably use methods and principles that have been developed for standard clinical trials, as well as for cluster randomized trials, with the additional caveat that we will always have to account for time and we will always have to account for the more complex correlation structure. So I think the two main methods of dealing with missing data, probably a complete case analysis, which is basically just analyzing all the available data, and multiple imputation, which is a process of randomly imputing missing values from a multi-variable model that's used to generate a distribution of plausible values.

Monica Taljaard: 16:12 So multiple imputation is the gold standard method, but it's complicated and quite frankly it's not always required in a randomized controlled trial. So, for example, if the outcome variable is missing in a step wedge design, here, I think the best advice might be to just analyze all the data that you have available. But you should adjust in the analysis for covariates that you think might help explain the missing data. You don't really need to use multiple imputation for missing outcomes because a complete case analysis is perfectly valid under this assumption of missing at random, which means we assume that the probability of having a missing outcome depends only on the observed covariates, as long as those covariates are also adjusted for in our analysis. But of course, it's very difficult to prove that this assumption of missing at random applies. So I think why... Sorry.

Vince Mor: 17:12 Let me just go one step further with that. So if I understand correctly then really the problem since the number of observations or the number of outcomes measured within any step in the step wedge design is probably less important for the ultimate power of of your efficiency of your design. Then really it's, you're assuming that the process that generated the missing notice of the outcome variable is the same across all of the wedges because that's sort of like assuming the

		experimentals, the controls, are subject to the same kind of bias. So is it that more than just a assumption of missing this at random?
Monica Taljaard:	17:59	Yes, for sure. So to get unbiased estimates of your treatment effect, you will have to assume that the missing mechanism or whatever factors drive this attrition or the missingness is non-differential across time and across the treatment conditions. In other words, that it's not an issue of the individuals being exposed to the intervention and therefore, they are less likely to want to complete the questionnaire, as that would be a really bad situation to be in. So you are hoping that the reasons why people are not completing the questionnaires could be just that there are some other things going on that potentially are captured with the baseline covariates. Maybe it's based on the baseline level of co-morbidity or some other factors that you've actually observed already at baseline. And as long as you adjust for those factors, you can still get unbiased estimates of the intervention effect.
Vince Mor:	19:04	Well, thank you very much. I actually, I've learned a huge amount and I have several follow up questions I'll ask you separately, but thank you very much for your time, and for that great talk.
Monica Taljaard:	19:15	Well, you're very welcome. Vince, it's been a great pleasure to do this.
Jill Harrison:	19:22	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.