What Is Outcome Clustering, and How Can It Affect a Trial?

Liz Turner:

So we randomize at the level of the clinic. This is the cluster. So we see that level two, the intervention in this case, is actually directly applied at the level of the clinic. And then we measure outcomes in Level 1 on individuals. So patient level outcomes, which are nested within the clinics. And this is specifically, did individuals agree to be screened for colorectal cancer? Now, importantly in the case of this trial, there are other factors related to the case of screening, but thanks to randomization, we should really be able to break the link between the intervention and those factors. We are more concerned about the potential for baseline imbalance in cluster randomized trials because we often enroll on not a lot of clusters.

So these individual level outcomes in this case, whether or not patients screened for colorectal cancer or not, these individual level outcomes within the same clinic are expected to be correlated, perhaps cause of the patient pool being similar because of the way behaviors occur. It could be a whole host of reasons.

In this case, what plays out is this is expected to reduce the power to detect a treatment effect if the same sample size would be used in a regular RCT, so an individually randomized control trial.

We want to try demystifying notions of what clustering is. So very importantly, we're really talking about outcome clustering. So in the case of the colorectal cancer screening trial, whether or not an individual agreed to be screened or not. So we have this binary outcome. Let's say it's refused screening. Yes, no. Let's consider 10 control arm clinics, each with five age eligible patients. So those who are not up-to-date with their colorectal cancer screening. And let's look at a range of different scenarios of different degrees of clustering.

Here's the first. We have our 10 clinics with five patients, age-eligible patients, in each of those clinics. And there are some clinics where all the patients are not screened, they're all red. And there are four clinics in which each of the five patients has been screened. So that's purple. So we see a lot of structure here by clinic.

In fact, what we see is the information we obtained from more than one participant in a clinic gives no more information than had we had a single participant. Precisely because every participant, so every patient in a given clinic has the same outcome. So there's a lot of structure by clinic, there's no within clinic variability, and we essentially see what's called here complete clustering. Our effective sample size is not the 50 patients. In fact, we're really only getting information from 10 units, those 10 clinics because within each clinic, all patients have exactly the same outcome. They either all screen or they all do not screen.

Now let's see a case where there's no structure by clinic, what's known as no clustering. In fact, what we see, we see there's a single purple dot out of the five in each of these 10 clinics. What does this mean?

Well, overall, there's a 20% uptake of colorectal cancer screening in each clinic. And similarly there is a 20% uptake across all 10 clinics. There's no structure by clinic. What this means is it's actually much more like a random sample of eligible participants. So in fact, this is much more like having information from 50 participants rather than the 10 clinics where the clinics would then be the unit of randomization in a trial. And remember, just for a moment, I'm just imagining the control arm clinics to not bring in the complexity of an intervention effect.

So this is a case of no clustering. This is good for us, we don't have structure by clinic, and therefore we effectively have the sample size of 50. Our effective sample size now is only 10~~,~~ You'd anticipate a lot of power loss when you have a very high degree of clustering.

Typically, in real life we were going to have some degree of clustering in between those two extremes. For example, what we see is a more typical situation. The proportion of screens ranging between 0% where we have that upper right clinic with all five patients, red have not screened up to about 80% say in the top left where we see four out of five are purple dots. So those four out of five patients have screened. So we do see some structure by clinic, some within clinic variability and some degree of clustering in this case.