How Does Clustering Affect Power and Detectable Difference?

Liz Turner:

On the horizontal actions we have the number of patients per cluster ranging from zero to three 50 per cluster. So say per clinic. On the vertical axis, we have the detectable difference in standard deviations unit. So here we're thinking of a difference between the two treatment arms of a cluster randomized trial and for a continuous outcome measured in standard deviation units. And we're going to also look at the plots when we change the degree of clustering. In this case it's an ICC Intraclass Correlation Coefficient of 0.1 and a fixed significance level, so a fixed alpha and power.

I'm going to show you four lines on each of these figures that showed increasing number of clusters per arm. So here we have, imagine you had a total of four clusters in a trial, so two clusters per arm. And you see the detectable difference here in standard deviation units is of the order of about two. And with increasing number of patients per cluster, there's not a huge impact on being able to detect smaller detectable differences. Although naturally as the sample size increases with more patients, we do have ability to detect slightly smaller effect sizes. Incidentally, we don't recommend conducting a cluster randomized trial with four clusters, we're using this just for the purposes of making a point on power.

So now we see we've doubled the sample size. We have now a total number of clusters of eight four in each arm of the trial and we've massively reduced the magnitude of effect. We can detect for a fixed degree of power. Conversely, we could also say in order to detect the original effect sizes of the order of two, we would have much greater power where we would have eight clusters in total rather than four. And then we see as we keep adding more clusters, we get our ability to detect even smaller effect sizes at the same level of power, the same significance level. And we see really the big impact of increasing the number of clusters is more important than increasing the number of patients per cluster.

Now we've taken that to be a 10th of the size ICC 0.01. So now we've seen, we can greatly reduce the effect sizes we can detect. So that's good news for us. We can detect much smaller signals now when we have a much lower degree of clustering because our effective sample size is not being penalized as much as it would've been had we had more clustering. So a larger degree of clustering measured by the ICC of 0.1 here, it's 0.01.

We see when we've reduced it even further to 0.001. So you see that we are slightly able to reduce the magnitudes of effects we can detect. So that's working in our favor. And in this case with the ICC of about 0.001, this is getting much closer to an individually randomized trial.

So clustering is kind of bad for us in many ways, we know it has a negative impact on power or equivalently, we can only detect larger effect sizes. We don't have ability to detect such small effect sizes when we have more clustering. But nevertheless, there's still many scenarios where cluster randomization is really the only appropriate choice. So we're going to have to design for it and account for it in our analysis.

So we're effectively going to inflate an RCT sample size. We definitely recommend working with a statistician as someone knowledgeable to do this.