Questions and Answers

Mind the Gap — Implications of Informative Cluster Size for the Design and Analysis of Cluster Randomized Trials

Brennan Kahan, Ph.D.

September 27, 2024

Q: How are the impacts of informative cluster size associated with the number of clusters in a trial? Are the effects of informative cluster even more pronounced when there are few clusters per arm?

A: In theory, the impacts of informative cluster size should be independent of the number of clusters in a trial, so it should have very similar impact in trials with few clusters as in trials with many clusters.

Q: Can IEEs lead to unstable estimates if weights are extreme or have high variance?

A: This generally shouldn't be a problem with IEEs, as they either weight by participants (for participant-average estimands) or by clusters (for cluster-average estimands), so there usually shouldn't be extreme weights. However, using different weighting schemes can affect the variance even if it doesn't lead to unstable estimates, so this may be something to consider at the design stage.

Q: How do we adjust for baseline covariates to account for ICS if we only have cluster level data?

A: If there is only cluster-level baseline data, this could be adjusted for in the model to try and reduce the impact of ICS. However, this will only work if the available cluster-level baseline data fully explains the ICS, which may not be a plausible assumption in practice.

Q: Issues pertaining to the number of clusters, variation in cluster sizes, and ICS are present in many studies. Should any of these issues be addressed before the others? Is there a general priority in which these issues should be handled?

A: There is not clear priority for the order in which these issues should be handled; they should all be carefully considered at the trial's design stage.

Q: Achieving no ICS is virtually impossible in real world trials. Is there any threshold for ICS below which we can ignore it when using GEE or mixed models?

A: The impact that ICS has on results from GEEs with an exchangeable correlation structure and mixed-effects models depends on a number of factors, including (a) the variation in cluster-size; (b) the magnitude of difference in outcomes and/or treatment effects between larger and smaller clusters; and (c) the ICC (the degree of within-cluster correlation). Because the impacts of ICS depend on these multiple factors, it is difficult to point to a specific threshold for safely using these methods when there is ICS.

Q: In a longitudinal trial where the “clusters” are individual people, there might be a situation where follow up measurements are missing for some people at some timepoints. So there are more measurements for some individuals than others (i.e. different cluster sizes). And people who are more engaged with the trial might have better outcomes. Is that an instance of informative cluster size? Should we avoid GEEs and mixed effects models in that sort of situation?

A: This would be akin to the informative cluster size setting if the decision to record a measurement were influenced by the patient’s health status in some way, for example if doctors had more frequent follow-ups for higher-risk patients. This may be frequent in studies using registry data or Electronic Health Records. However, this is different to the setting of a randomised trial where some patients miss their scheduled follow-up; this is a missing data problem, and here mixed-effects models will provide valid results so long as the assumptions about the missing data are correct (e.g. that data are missing-at-random).