HOW TO APPLY THE MULTIPHASE OPTIMIZATION STRATEGY (MOST) IN YOUR INTERVENTION DEVELOPMENT RESEARCH

Module 4
Some conceptual and technical aspects of the factorial experiment

Lesson 5: Recognizing and dealing with a cluster structure

This course was developed by

Linda M. Collins

School of Global Public Health

New York University

(narrator)

Kate Guastaferro

College of Health and Human Development

The Pennsylvania State University

In the previous lesson you learned how to:

- Distinguish between the conclusion-priority and decision-priority perspectives
- Discern whether the conclusion-priority or decision-priority perspective is appropriate in a given situation

In this lesson you will learn how to

- Recognize when a cluster structure is present
- Understand how a cluster structure can affect statistical power

What is a cluster structure in data?

- This is also called nesting
- A cluster structure occurs whenever participants are members of clusters, such as schools, neighborhoods, medical practices, and the like
- Sometimes participants within clusters tend to be more alike, or are expected to become more alike, than participants in the overall target population

Two different types of clusters: pre-existing and experiment-induced

- Examples of pre-existing clusters
  - Clinics or medical practices
  - Schools or classrooms
  - Families
  - Faith congregations
  - Communities

Experiment-induced clusters

- Clusters that are formed in the course of experimentation
- Example: Suppose one candidate component is a treatment that is delivered in a group setting
  - Participants become more similar over course of treatment because of group interaction
  - A concern even when participants randomly assigned to groups

The intra-class correlation

- Degree of within-cluster similarity in Y expressed in the intraclass correlation ICC_Y

Why is it important to be aware of clustering? Reason 1

- Contamination between experimental conditions
  - When treatment "bleeds" across experimental conditions
- Contamination tends to reduce treatment effects

Why is it important to be aware of clustering? Two main reasons. Reason 1

- Example:
  - Married couple both participants in smoking cessation experiment
  - One assigned to treatment in which nicotine lozenges are provided, other to control without any NRT
  - Spouse in treatment condition gives some of the lozenges to spouse in control condition

Why is it important to be aware of clustering? Reason 1

- Another example:
  - Consider an optimization trial in which one of the candidate components is a treatment delivered in a group setting with a lot of discussion
  - In addition, there are several other components being examined
  - People in the treatment group discuss which components they have been assigned to receive and share information

Why is it important to be aware of clustering? Reason 2

- The intervention may be aimed at both the individual and cluster level (or solely at the cluster level)
- Example: school-based drug abuse prevention delivered in classroom
  - One target: individual children
  - Another target: social norms about drug abuse at classroom and school level

Why is it important to be aware of clustering?

- If…
  - …there could be a substantial amount of contamination OR
  - …the intervention is aimed at both the individual and cluster level (or solely the cluster level)
- Then…
  - …it does not make sense for individuals within a cluster to be in different experimental conditions

Two types of random assignment

Within-cluster randomization

- Also called individual randomization
- An individual in a given cluster may be assigned to any experimental condition

Between-cluster randomization

- Also called cluster randomization
- Entire clusters assigned to experimental conditions, so all individuals in a cluster in same condition

Effect of clustering on statistical power: It depends on which type of random assignment

- If you use within-cluster random assignment, cluster structure has little effect on power
- If you use between-cluster random assignment, cluster structure can have a huge effect on power
  - Earlier in this module we alluded to another consideration related to power in addition to N, α, and effect size
  - We were referring to the effect of between-cluster random assignment as expressed in D

Effect of clustering on statistical power: It depends on which type of random assignment

- Impact of cluster structure expressed in the design effect D (Dziak et al., 2012; Murray, 1998)
- D is a multiplier; if you would need N without a cluster structure you need DN with a cluster structure
- Example: If power analysis indicates you need N=400, and D=1.2, you need 400 X 1.2 = 480.

Effect of clustering on statistical power: It depends on which type of random assignment

$$D = 1 + (n-1)ICC_X ICC_Y$$

ICC_X= intraclass correlation for the X variables (predictors – the codes representing main effects and interactions)

ICC_Y= intraclass correlation for Y variable (outcome)

n = number per cluster

Effect of clustering on statistical power: Within-cluster random assignment

$$D = 1 + (n - 1) ICC_X ICC_Y$$

- It turns out that with within-cluster randomization, if all participants have the same assignment probabilities, $ICC_X = 0$
- Therefore $D = 1$
- One way to see this: Knowing a participant's cluster membership reveals nothing about what condition they have been assigned to
- Conclusion: little to no effect on power

Effect of clustering on statistical power: Within-cluster random assignment

$$D = 1 + (n - 1) ICC_X ICC_Y$$

- How to ensure that all participants have the same assignment probabilities?
  - Be sure to replicate the entire experiment within each cluster

Effect of clustering on statistical power: Within-cluster random assignment

- Example: Suppose you are conducting a 2^4 experiment in 5 different clinics. You feel comfortable conducting within-cluster random assignment.
- Then you would implement the entire 2^4 experiment in each of the clinics
  - A participant in any clinic would have a 1/16 chance of being assigned to any one of the experimental conditions

Effect of clustering on statistical power: Within-cluster random assignment

- Note we are talking just about power here
- There may be other considerations
  - e.g. variability in effects across clusters

Effect of clustering on statistical power: Between-cluster random assignment

$$D = 1 + (n - 1)ICC_X ICC_Y$$

- Here clusters are randomized to experimental conditions
- Then cluster membership determines experimental condition
- Therefore $ICC_X = 1$, so

$$D = 1 + (n - 1)\,ICC_Y$$

Effect of clustering on statistical power: Between-cluster random assignment

$$D = 1 + (n - 1)\, ICC_Y$$

- You can see that all else being equal, D is larger when
  - Cluster size n is larger
  - ICC_Y is larger

Factorial optimization trials when between-cluster randomization is necessary: Two considerations

- Will I have enough power?
- Issue essentially the same as it would be for an RCT
- Described in Murray’s (1998) classic book

Factorial optimization trials when between-cluster randomization is necessary: Two considerations

- Given that I have enough power, will I have enough clusters to assign to experimental conditions?
- Good idea to assign at least 2 clusters/condition
- May want to consider a fractional factorial design (see Dziak et al., 2012)

For additional information

- For more about experiment-induced clustering in factorial optimization trials, see Nahum-Shani & Dziak (2018) and Nahum-Shani et al. (2018)
- For clustering and SMARTs, see NeCamp et al. (2017)
- This is still an open research area

In this lesson you learned how to

- Recognize when a cluster structure is present
- Understand how a cluster structure can affect statistical power

In the next lesson you will

- Review what you have learned in Module 4

References Cited

- Dziak, J. J., Nahum-Shani, I., & Collins, L. M. (2012). Multilevel factorial experiments for developing behavioral interventions: power, sample size, and resource considerations. *Psychological Methods*, *17*, 153-175.
- Murray, D. M. (1998). *Design and analysis of group-randomized trials*. Oxford: Oxford University Press.
- Nahum-Shani, I., & Dziak, J.J. (2018). Multilevel factorial designs in intervention development. In L.M. Collins and K.C. Kugler (eds.), *The multiphase optimization strategy (MOST): Advanced topics*. New York: Springer.
- Nahum-Shani, I., Dziak, J. J., & Collins, L. M. (2018). Multilevel factorial designs with experiment-induced clustering. *Psychological Methods*, *23*, 458-479.
- NeCamp, T., Kilbourne, A., & Almirall, D. (2017). Comparing cluster-level dynamic treatment regimens using sequential, multiple assignment, randomized trials: Regression estimation and sample size considerations. *Statistical methods in medical research*, *26*, 1572-1589.