If you're a person with a disability or using assistive technology and are having difficulty accessing any of the content in this file, please contact the ODP at prevention@nih.gov

Methods: Mind the Gap
Webinar Series

Design and Analytic Methods for Group-Based Interventions

Presented by:

David Murray, Ph.D.
Associate Director for Prevention,
Director, Office of Disease Prevention
National Institutes of Health

National Institutes of Health
Office of Disease Prevention

35 YEARS

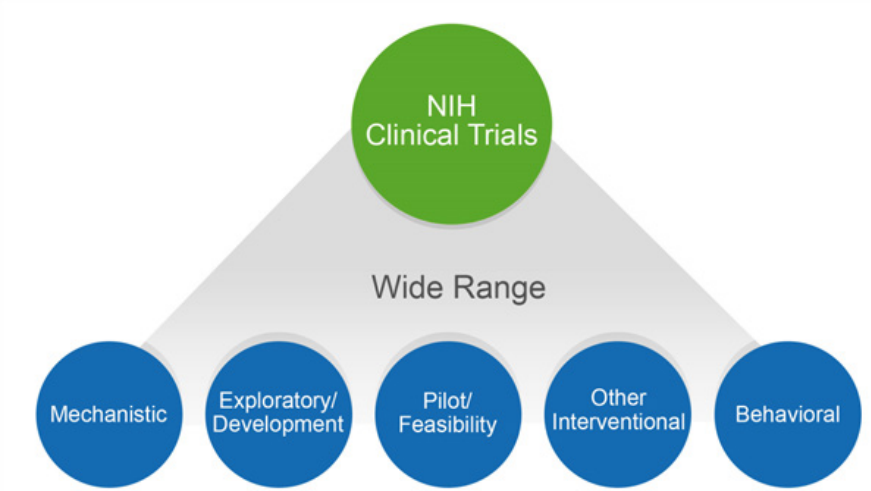
Office of Disease Prevention
Putting Prevention Research First

Agenda

- NIH definition of a clinical trial
- Group- or cluster-randomized trials (GRTs)
- Individually randomized group-treatment trials (IRGTs)
- Research Methods Resources website

NIH Definition of a Clinical Trial

- NIH published its revised definition of a clinical trial in 2014 after extensive public input.
  - A research study in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or behavioral outcomes.



Does Your Study Meet the Definition?

- Does the study involve human participants?
- Are the participants prospectively assigned to an intervention?
- Is the study designed to evaluate the effect of the intervention on the participants?
- Is the effect that will be evaluated a health-related biomedical or behavioral outcome?

- If the answer to all four questions is yes, your study is a clinical trial under the NIH definition.

- Information on the Clinical Trial Requirements for Grants and Contracts is available at:
  - https://grants.nih.gov/policy/clinical-trials.htm

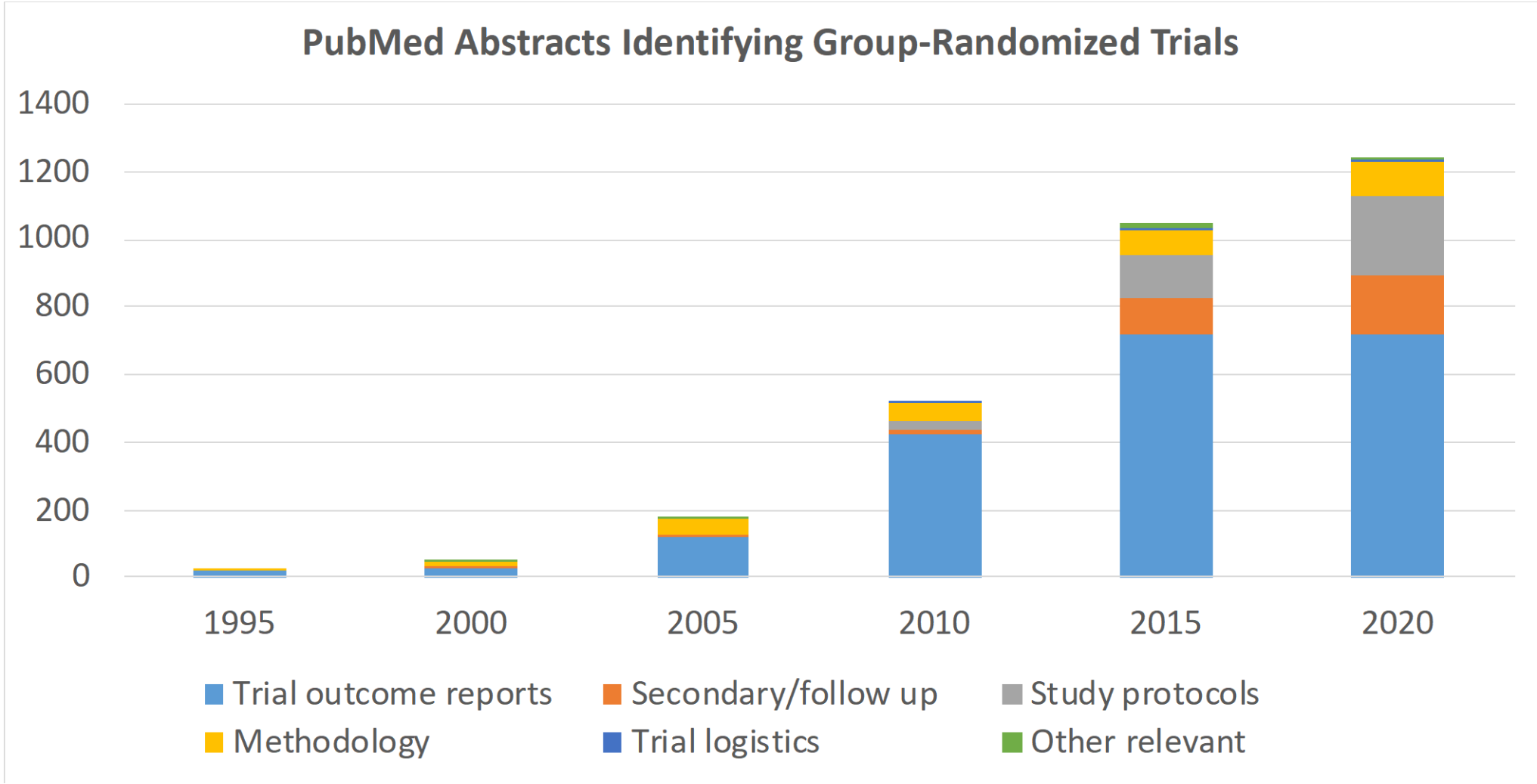
Three Kinds of Randomized Trials

- Randomized Clinical Trials (RCTs)
  - Individuals randomized to study conditions with no interaction among participants after randomization (no group sessions, virtual interaction, or shared intervention agent)
    - Most drug trials
- Individually Randomized Group Treatment Trials (IRGTs)
  - Individuals randomized to study conditions with interaction among participants after randomization or with a shared intervention agent
    - Many surgical trials
    - Many behavioral trials
- Group-Randomized Trials (GRTs)
  - Groups randomized to study conditions with interaction among the members of the same group before and after randomization
    - Many trials conducted in communities, worksites, schools, clinics, etc.
    - We will focus on parallel GRTs, ignoring cross-over and stepped wedge designs.

Two Kinds of Group-Randomized Trials

- Parallel GRT
  - Separate but parallel intervention and control conditions throughout the trial, with no crossover.
- Stepped Wedge GRT
  - All groups start in the control condition.
  - All groups crossover to the intervention condition, but in a random order and on a staggered schedule.
  - All groups receive the intervention before the end of the study.

PubMed Abstracts Identifying GRTs



Impact on the Design

- Randomized clinical trials
  - There is usually good opportunity for randomization to distribute potential confounders evenly, as most RCTS have N>100.
  - If well executed, confounding is not usually a concern.
- Individually randomized group treatment trials
  - There may be less opportunity for randomization to distribute potential confounders evenly, as many IRGTs have N<100.
  - Confounding can be more of a concern in IRGTs than in RCTs.
- Parallel group-randomized trials
  - GRTs often involve a limited number of groups, often <50.
  - In any single realization, there is limited opportunity for randomization to distribute all potential confounders evenly.
  - Confounding is a concern in GRTs if G<50.

Impact on the Analysis in a GRT or IRGT

- Observations on randomized individuals who do not interact are independent and are analyzed with standard methods.
- The members of the same group in a GRT will share some physical, geographic, social or other connection.
- The members of groups in an IRGT will develop similar connections.
- Those connections will create a positive intraclass correlation that reflects extra variation attributable to the group.

$$ICC_{m:g:c} = corr\left(y_{i:k:l}, y_{i':k:l}\right)$$

- The positive ICC reduces the variation among the members of the same group so the within-group variance is:

$$\sigma_e^2 = \sigma_y^2\left(1 - ICC_{m:g:c}\right)$$

Impact on the Analysis in a GRT or IRGT

- The between-group component is the one's complement:

$$\sigma^2_{g:c} = \sigma^2_y \left(ICC_{m:g:c} \right)$$

- The total variance is the sum of the two components:

$$\sigma^2_y = \sigma^2_e + \sigma^2_{g:c}$$

- The intraclass correlation is the fraction of the total variation in the data that is attributable to the unit of assignment:

$$ICC_{m:g:c} = \frac{\sigma^2_{g:c}}{\sigma^2_e + \sigma^2_{g:c}}$$

Impact on the Analysis in a GRT or IRGT

- Given m members in each of g groups...
  - When group membership is established by random assignment,

$$\sigma^2_{\bar{y}_g} = \frac{\sigma^2_y}{m}$$

  - When group membership is not established by random assignment,

$$\sigma^2_{\bar{y}_g} = \frac{\sigma^2_e}{m} + \sigma^2_g$$

  - Or equivalently,

$$\sigma^2_{\bar{y}_g} = \frac{\sigma^2_y}{m}\left(1 + (m-1)\,ICC\right)$$

Impact on the Analysis in a GRT or IRGT

- Nested factors must be modeled as random effects (Zucker, 1990).
- The variance of any group-level statistic will be larger.
- The df to estimate the group-level component of variance will be based on the number of groups, and so is often limited.
  - This is almost always true in a GRT, can be true in an IRGT.
- Any analysis that ignores the extra variation or the limited df will have a Type I error rate that is inflated, often badly.
  - Type I error rate may be 30-50% in a GRT, even with small ICC
  - Type I error rate may be 15-25% in an IRGT, even with small ICC
- Extra variation and limited df always reduce power.

- Zucker DM. An analysis of variance pitfall: The fixed effects analysis in a nested design. Educ and Psych Measurement. 1990;50(4):731-8.

The Warning

Randomization by cluster accompanied by an analysis appropriate to randomization by individual is an exercise in self-deception, however, and should be discouraged.

Cornfield (1978)

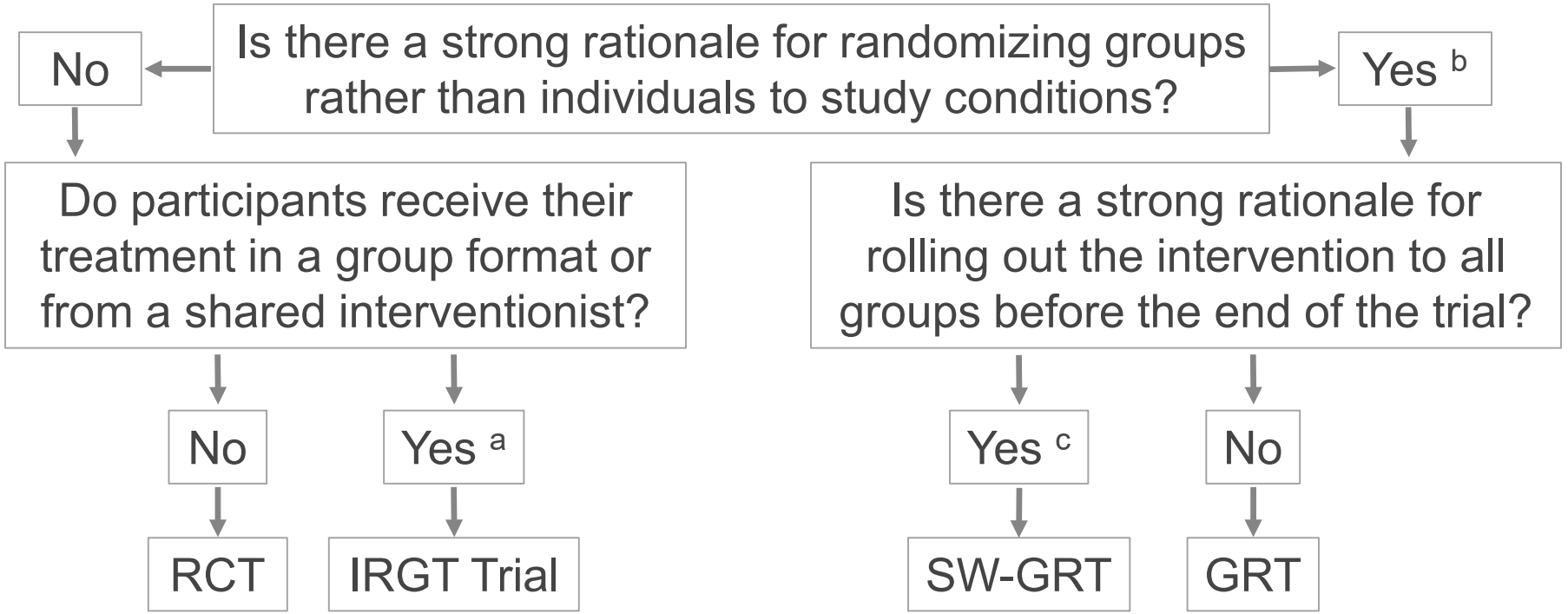
- Though Cornfield's remarks were addressed only to GRTs, they also apply to IRGTs

- Cornfield J. Randomization by group: a formal analysis. Am J Epi. 1978;108(2):100-2.

The Need for GRTs and IRGTs

- An RCT is the best comparative design when individual randomization is possible without post-randomization interaction.
- An IRGT is the best comparative design whenever...
  - Individual randomization is possible but there are good reasons to deliver the intervention in a group format or through a shared interventionist
- A GRT is the best comparative design whenever the investigator wants to evaluate an intervention that…
  - Manipulates the social or physical environment or cannot be delivered to individuals without risk of contamination

Choosing Among These Designs



[a] If the intervention is delivered through a physical or a virtual group, or through shared interventionists who each work with multiple participants, positive ICC can develop over the course of the trial.

[b] There may be logistical reasons to randomize groups or it may not be possible to deliver the intervention to individuals without substantial risk of contamination.

[c] There may be good political or logistical reasons to roll out the intervention to all groups before the end of the trial.

Adapted from Murray DM, Taljaard M, Turner EL, George SM. Essential Ingredients and Innovations in the Design and Analysis of Group-Randomized Trials. Annual Review of Public Health. 2020;41:1-19. PMID31869281.

Preferred Analytic Models for Parallel GRT Designs With One or Two Time Intervals

- Mixed-model ANOVA/ANCOVA
  - Extension of the familiar ANOVA/ANCOVA based on the General Linear Model
  - Fit using the General Linear Mixed Model or the Generalized Linear Mixed Model
  - Accommodates regression adjustment for covariates
  - Can not misrepresent over-time correlation
  - Can take several forms
    - Posttest-only ANOVA/ANCOVA
    - ANCOVA of posttest with regression adjustment for pretest
    - Repeated measures ANOVA/ANCOVA for pretest-posttest design
  - Simulations have shown these methods have the nominal Type I error rate across a wide range of conditions common in GRTs.

- Murray DM. Design and Analysis of Group-Randomized Trials. New York, NY: Oxford University Press; 1998.
- Donner A, Klar N. Design and Analysis of Cluster Randomization Trials in Health Research. London: Arnold; 2000.

Preferred Analytic Models for Parallel GRT Designs With More Than Two Time Intervals

- Random coefficients models
  - Also called growth curve models
  - The intervention effect is estimated as the difference in the condition mean trends.
  - Mixed-model ANOVA/ANCOVA assumes homogeneity of group-specific trends.
    - Simulations have shown that mixed-model ANOVA/ANCOVA has an inflated Type I error rate if those trends are heterogeneous (Murray et al., 1998).
    - An unstructured covariance matrix will not avoid this problem (unpublished).
  - Random coefficients models allow for heterogeneity of those trends.
  - Simulations have shown these methods have the nominal Type I error rate across a wide range of conditions common in GRTs.

- Murray DM, Hannan PJ, et al. Analysis of data from group-randomized trials with repeat observations on the same groups. Stat Med. 1998;17(14):1581-600. PMID9699231.

Preferred Analytic Models for Individually Randomized Group Treatment Trials

- Analyses that ignore the ICC risk an inflated Type I error rate (cf. Pals et al., 2008; Baldwin et al., 2011).
  - Not as severe as in a GRT but can exceed 15% under conditions common to these studies.
  - The solution is the same as in a GRT.
    - Analyze to reflect the variation attributable to the groups defined by the patterns of interaction.
    - Base df on the number of groups, not the number of members.
  - Mixed models are the most common approach.
    - Candlish et al. (2018) provide SAS, R, and STATA code for analyzing IRGT trials.

- Pals SL, Murray DM, et al. Individually randomized group treatment trials: a critical appraisal of frequently used design and analytic approaches. Am J Public Health. 2008;98(8):1418-24. PMID18556603.
- Baldwin SA, Bauer DJ, et al. Evaluating models for partially clustered designs. Psych Methods. 2011;16(2):149-65. PMID21517179.
- Candlish J et al. Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: a simulation study. BMC Med Res Methodol. 2018;18(1):105. PMID: 30314463.

IRGT Trials: Cross-Classification, Multiple Membership, or Dynamic Groups

- The GRT and IRGT literature assumes that each member belongs to one group and that group membership does not change over time.
  - These patterns often do not hold in practice and failure to model the correct structure can lead to an inflated type 1 error rate.
  - Roberts and Walwyn (2013), Luo et al. (2015), and Sterba (2017) describe cross-classified, multiple membership, and dynamic group models that address these complex design features.

- Roberts C, Walwyn R. Design and analysis of non-pharmacological treatment trials with multiple therapists per patient. Stat Med. 2013;32(1):81-98. PMID22865729.
- Luo W, Cappaert KJ, et al. Modelling partially cross-classified multilevel data. Br J Math Stat Psychol. 2015;68(2):342-62. PMID25773173.
- Sterba SK. Partially nested designs in psychotherapy trials: A review of modeling developments. Psychother Res. 2017;27(4):425-36. PMID26686878.

Factors That Affect Precision in a Parallel GRT

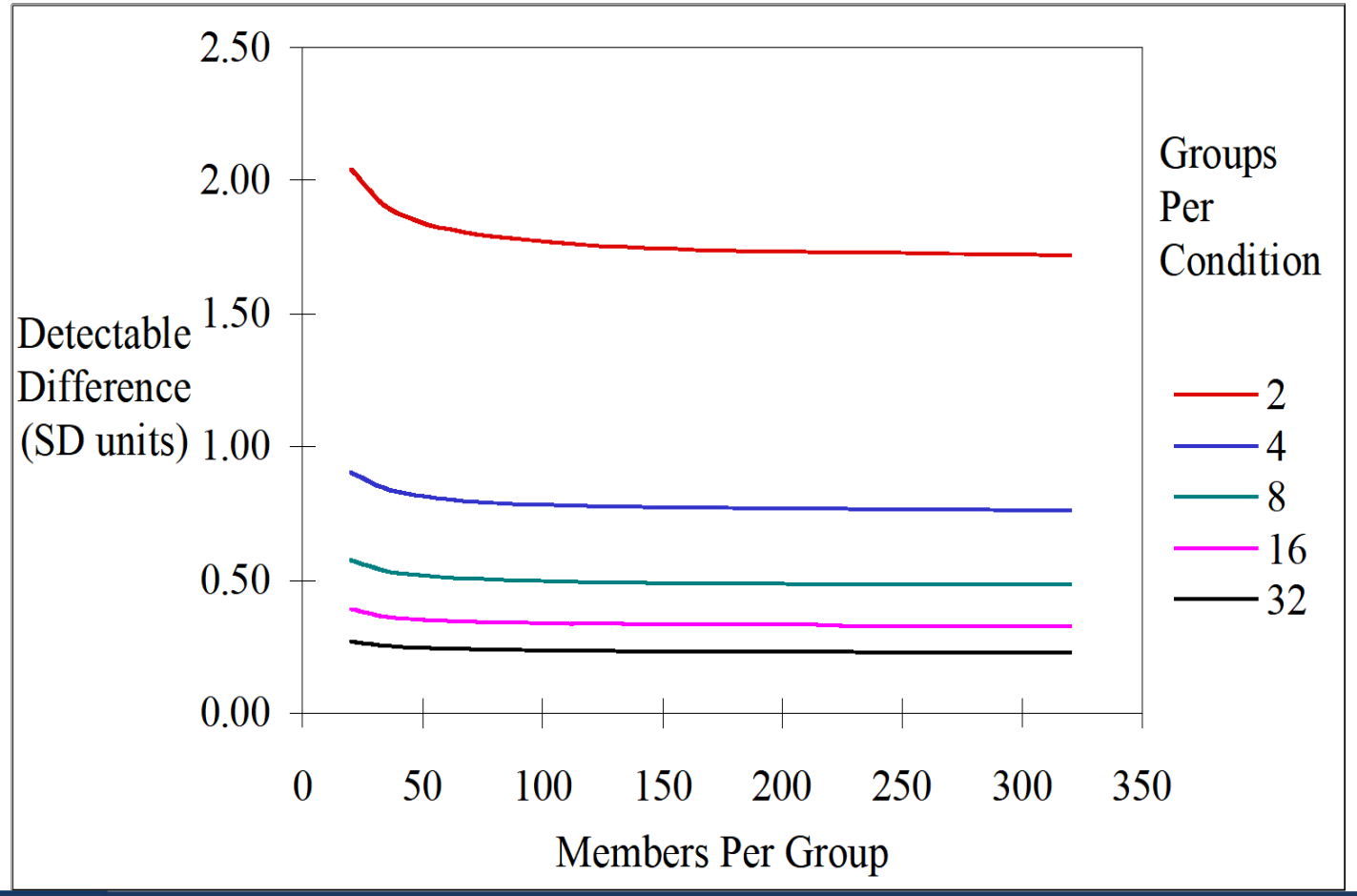
- The variance of the condition mean in a parallel GRT is:

$$\sigma^2_{\bar{y}_c} = \frac{\sigma^2_y}{mg}\left(1 + (m-1)ICC\right)$$

- This equation must be adapted for more complex analyses, but the precision of the analysis will always be directly related to the components of this formula operative in the proposed analysis:
  - Replication of members and groups
  - Variation in measures
  - Intraclass correlation

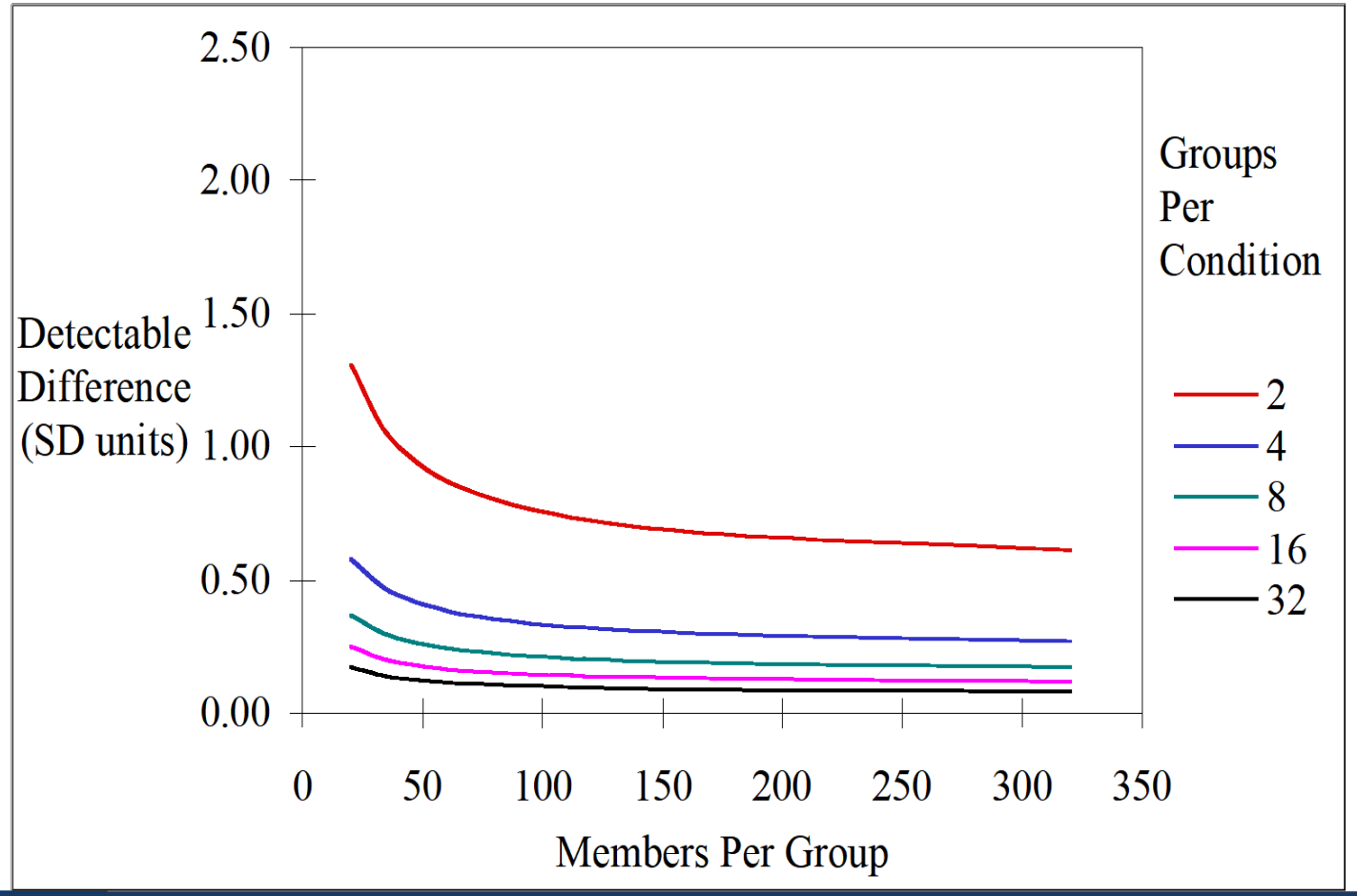
Improving Precision in a Parallel GRT

- Increased replication (ICC=0.100)



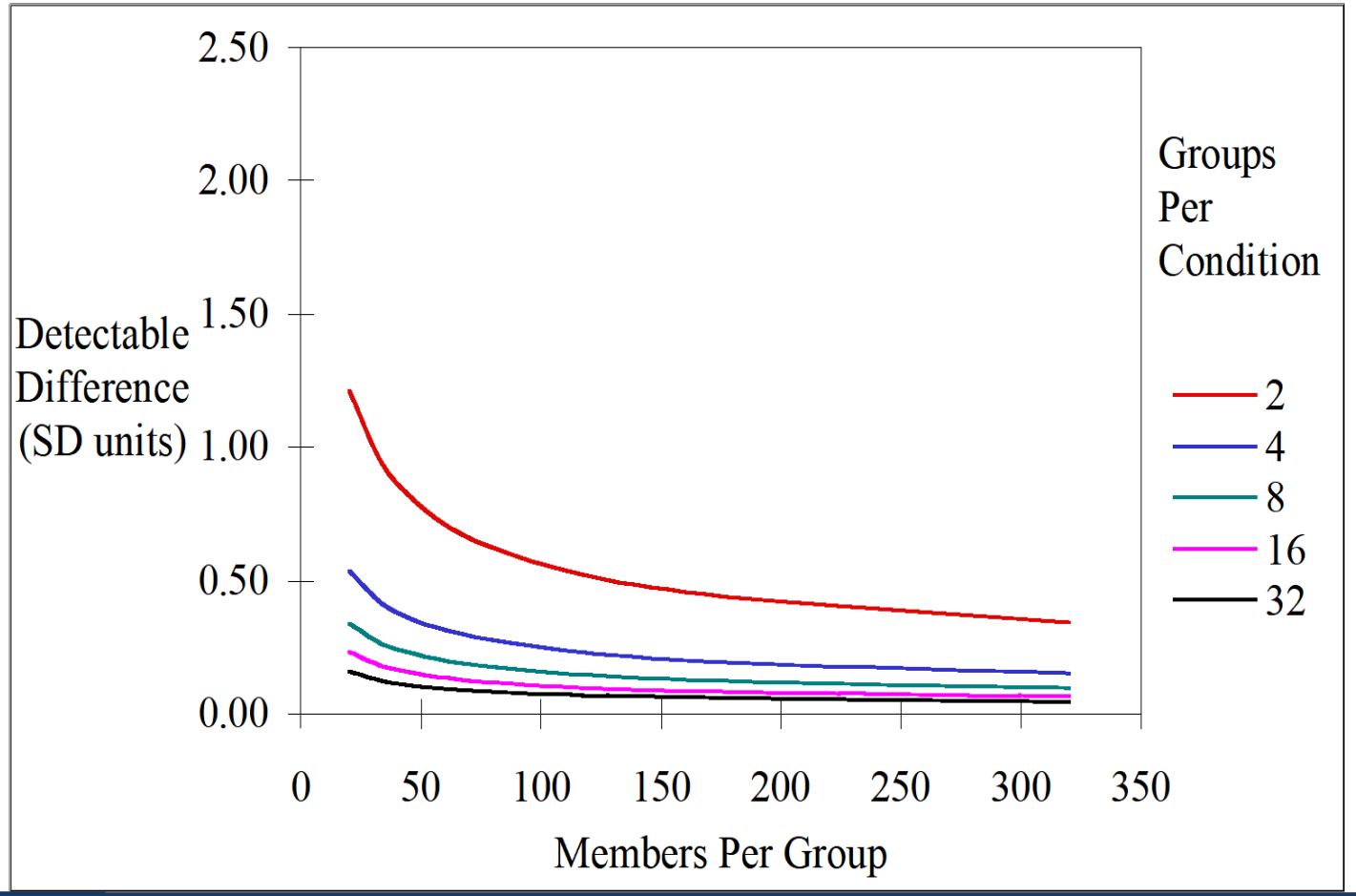
Improving Precision in a Parallel GRT

- Reduced ICC (ICC=0.010)



Improving Precision in a Parallel GRT

- The law of diminishing returns (ICC=0.001)



Power for Parallel GRTs

- The usual methods must be adapted to reflect the nested design
  - The variance is greater in a parallel GRT due to the expected ICC.
  - df should be based on the number of groups, not the number of members.
- Many papers now report ICCs and show how to plan a parallel GRT.
- Power in parallel GRTs is tricky, and investigators are advised to get help from someone familiar with these methods.
- A good resource is the NIH Research Methods Resources website
  - https://researchmethodsresources.nih.gov

Power for IRGTs

- Power depends heavily on the ICC, the number of groups per condition, and the number of members in the control condition for IRGTs with groups in one condition.
- Power is better in trials that do not have post-randomization interaction in the control condition.
- Methods for sample size estimation for IRGTs have been published.
  - Pals SL, Murray DM et al. Individually randomized group treatment trials: a critical appraisal of frequently used design and analytic approaches. Am J Pub Health. 2008;98(8):1418-24. PMID18556603.
  - Roberts C, Walwyn R. Design and analysis of non-pharmacological treatment trials with multiple therapists per patient. Stat Med. 2013;32(1):81-98. PMID22865729.
  - Moerbeek M, Teerenstra S. Power analysis of trials with multilevel data. Boca Raton: CRC Press; 2016.
  - Hemming K, Kasza J, et al. A tutorial on sample size calculation for multiple-period cluster randomised parallel, cross-over and stepped-wedge trials using the Shiny CRT Calculator. Int J of Epidem. 2020;49(3):979-95. PMID32087011.
- The NIH Research Methods Resources website added an IRGT sample size calculator: https://researchmethodsresources.nih.gov

Unbalanced Designs

- Most of the methods for sample size estimation and data analysis assume a balanced design in terms of group size.
- As long as the ratio of the largest to the group is no worse than about 2:1, those methods are fine.
- Given more extreme imbalance reduces power and can lead to an inflated type I error rate if ignored in the analysis.

- Candel MJ, Van Breukelen GJ. Varying cluster sizes in trials with clusters in one treatment arm: sample size adjustments when testing treatment effects with linear mixed models. Stat Med. 2009;28(18):2307-24.
- Candel MJ, Van Breukelen GJ. Sample size adjustments for varying cluster sizes in cluster randomized trials with binary outcomes analyzed with second-order PQL mixed logistic regression. Stat Med. 2010;29(14):1488-501.
- You Z, Williams OD, Aban I, Kabagambe EK, Tiwari HK, Cutter G. Relative efficiency and sample size for cluster randomized trials with variable cluster sizes. Clinical Trials. 2011;8(1):27-36.
- Candel MJ, Van Breukelen GJ. Repairing the efficiency loss due to varying cluster sizes in two-level two-armed randomized trials with heterogeneous clustering. Stat Med. 2016;35(12):2000-15.
- Moerbeek M, Teerenstra S. Power analysis of trials with multilevel data. Boca Raton: CRC Press; 2016.
- Hemming K, Kasza J, et al. A tutorial on sample size calculation for multiple-period cluster randomised parallel, cross-over and stepped-wedge trials using the Shiny CRT Calculator. International Journal of Epidemiology. in press.

NIH Resources

- Pragmatic and Group-Randomized Trials in Public Health and Medicine
  - https://prevention.nih.gov/grt
  - 7-part online course on GRTs and IRGTs
- Mind the Gap Webinars
  - https://prevention.nih.gov/education-training/methods-mind-gap
    - Overview of Statistical Models for the Design and Analysis of Stepped Wedge Cluster Randomized Trials (Fan Li, July 14, 2020)
    - SW-GRTs for Disease Prevention Research (Monica Taljaard, July 11, 2018)
    - Design and Analysis of IRGTs in Public Health (Sherri Pals, April 24, 2017)
    - Research Methods Resources for Clinical Trials Involving Groups or Clusters (David Murray, December 13, 2017)
- Research Methods Resources Website
  - https://researchmethodsresources.nih.gov/
  - Material and sample size calculators for parallel GRTs and IRGTs.

National Institutes of Health U.S. Department of Health and Human Services

NIH National Institutes of Health
Turning Discovery Into Health

Search

Contact Us

Research Methods Resources

Home | Methods | Tools | Reference Materials

The Research Methods Resources website provides investigators with important research methods resources to help them satisfy the NIH's clinical research requirements for trials that randomize groups or deliver interventions to groups.

Introductory Webinar

Trials that Randomize Groups or Deliver Interventions to Groups

Experiments, including clinical trials, differ considerably in the methods used to assign participants to study conditions (or study arms) and to deliver interventions to those participants.

This website provides information related to the design and analysis of experiments in which (1) participants are assigned in groups (or clusters) and individual observations are analyzed to evaluate the effect of the intervention, and (2) participants are assigned individually but receive at least some of their intervention with other participants or through an intervention agent shared with other participants.

This material is relevant for both human and animal studies as well as basic and applied research. And while it is important for investigators to become familiar with the issues presented on this website, it is even more important that they collaborate with a methodologist who is familiar with these issues.

Group-or Cluster-Randomized Trials (GRTs)

Individually Randomized Group-Treatment (IRGT) Trials

GRT Sample Size Calculator

Use this calculator to estimate sample sizes for GRTs.

There are nine steps for each sample size calculation. You will be asked to specify the type I error rate and desired power for the test of your intervention effect, the expected distribution of your outcome variable, and the design and analytic plan for your trial. You will also be asked to provide estimates of the parameters needed to estimate the sample size required for your study. Please note that this sample size calculator assumes the same variance components, ICCs, and group sizes in both study conditions; those assumptions are appropriate for most GRTs.

IRGT Sample Size Calculator

Use this calculator to estimate sample size for an IRGT trial with clustering in only one arm. If you expect to have clustering in both arms, you can use the GRT Sample Size Calculator, because GRTs always have clustering in both arms. Please note however, that the GRT Sample Size Calculator assumes the same variance components, ICCs, and group sizes across study conditions, which may not always be appropriate for an IRGT with clustering in both arms (Roberts and Roberts, 2005).

There are eight steps for each sample size calculation for an IRGT trial. You will be asked to specify the type I error rate and desired power for the test of your intervention effect, the expected distribution of your outcome variable, and the

GRT Sample Size Calculator

View Worked Examples

1. **Type I Error Rate and Power**
2. Expected Distribution of the Primary Outcome
3. Design and Analytic Plan
4. Intraclass Correlation
5. Members or Participants
6. Regression Adjustment for Covariates
7. Stratification and Matching
8. Analysis
9. Results

You will need to select the type I error rate and desired power for the test of the intervention effect. The type I error rate is the probability of making a type I error, which is to incorrectly reject the null hypothesis of no intervention effect. Generally, investigators try to limit that probability, so that they do not report chance findings as though they were real intervention effects. Power is the probability of correctly rejecting a false null hypothesis. In general, investigators try to maximize that

MORE

Enter the type I error rate for the two-tailed test of the intervention effect; if you plan a one-tailed test, double that value.

0.05

Min. 0.0001, Max. 0.2

IRGT Sample Size Calculator

View Worked Examples

1. **Type I Error Rate and Power**
2. Expected Distribution of the Primary Outcome
3. Design and Analytic Plan
4. Intraclass Correlation
5. Members or Participants
6. Regression Adjustment for Covariates
7. Analysis
8. Results

You will need to select the type I error rate and desired power for the test of the intervention effect. The type I error rate is the probability of making a type I error, which is to incorrectly reject the null hypothesis of no intervention effect. Generally, investigators try to limit that probability, so that they do not report chance findings as though they were real intervention effects. Power is the probability of correctly rejecting a false null hypothesis. In general, investigators try to maximize that

MORE

Enter the type I error rate for the two-tailed test of the intervention effect; if you plan a one-tailed test, double that value.

0.05

Min. 0.0001, Max. 0.2

Worked Example for An Analysis of a Simple Difference

The calculator will use the parameter estimates you have provided to calculate power for the intervention effect (Δ) that you specified.

In addition, the calculator will use the other parameter estimates to calculate the detectable difference (Δ) that is available with the level of power that you specified as a function of the number of groups in the intervention condition (g), the number of members in the control condition (m_c), and the number of members in each group in the intervention condition (m_i), using the formula below.

$$\Delta = \sqrt{\left(\left(\frac{\sigma_y^2\left(1-R_{y\cdot x_m}^2\right)}{m_c}\right)+\left(\frac{\left(\sigma_y^2(1-ICC)\left(1-R_{y\cdot x_m}^2\right)+\left(m_i\sigma_y^2(ICC)\left(1-R_{y\cdot x_g}^2\right)\right)\right)}{m_i g}\right)\right)\left(t_{\alpha/2}+t_{\beta}\right)^2}$$

Here σ_y^2 is the variance of the outcome variable ignoring any expected ICC, $R_{y\cdot x_m}^2$ is the proportion of variance explained by the member-level covariates, m is the number of

8.50 x 11.00 in

As an example, consider the following set of parameter estimates:

Let m_c = 50, m_i = 10, g = 5, with x_m=4 df used for member-level covariates and x_g=1 df used for group-level covariates. Then

$$df = (m_c - 1) + (g - 1) - df_x = (50 - 1) + (5 - 1) - 4 - 1 = 48$$

For a two-tailed type 1 error rate of 5%,

$$t_{\alpha/2} = 2.0106 \text{ and } t_{\beta} = 0.8492$$

Let $\sigma^2_{y_c}$ = 1.0 and ICC=0.05.

Let $R^2_{y \cdot x_m} = 0.30$ and $R^2_{y \cdot x_g} = 0.10$.

Then

$$\Delta = \sqrt{\left(\left(\frac{1\left(1-0.30\right)}{50}\right)+\left(\frac{\left(1.00\left(1-0.05\right)\left(1-0.30\right)+\left(10*1.00*0.05\left(1-0.10\right)\right)\right)}{10*5}\right)\right)\left(2.0106+0.8492\right)^2}$$

$$= 0.5449$$

IRGT Sample Size Calculator

View Worked Examples

1. Type I Error Rate and Power
2. Expected Distribution of the Primary Outcome
3. Design and Analytic Plan
4. Intraclass Correlation
5. Members or Participants
6. Regression Adjustment for Covariates
7. Analysis
8. Results

Reset and Start Over

Results

Power for the specified intervention effect based on the parameter estimates: **0.1707**

The power for the specified intervention effect based on your parameter estimates is less than the desired power of 0.80. Consider increasing the number of groups in the intervention condition, the number of members in each group in the intervention condition, or the number of members in the control condition. Be cautious about reducing the ICC, or increasing the over-time correlations or the proportion of variance explained by covariates, as those should be data-based estimates and not changed simply because the power is insufficient.

Parameter Estimates

Type I Error Rate and Power

Enter the type I error rate for the test of the intervention effect: **0.05**

Enter the desired power for the test of the intervention effect: **0.80**

Expected Distribution of the Primary Outcome

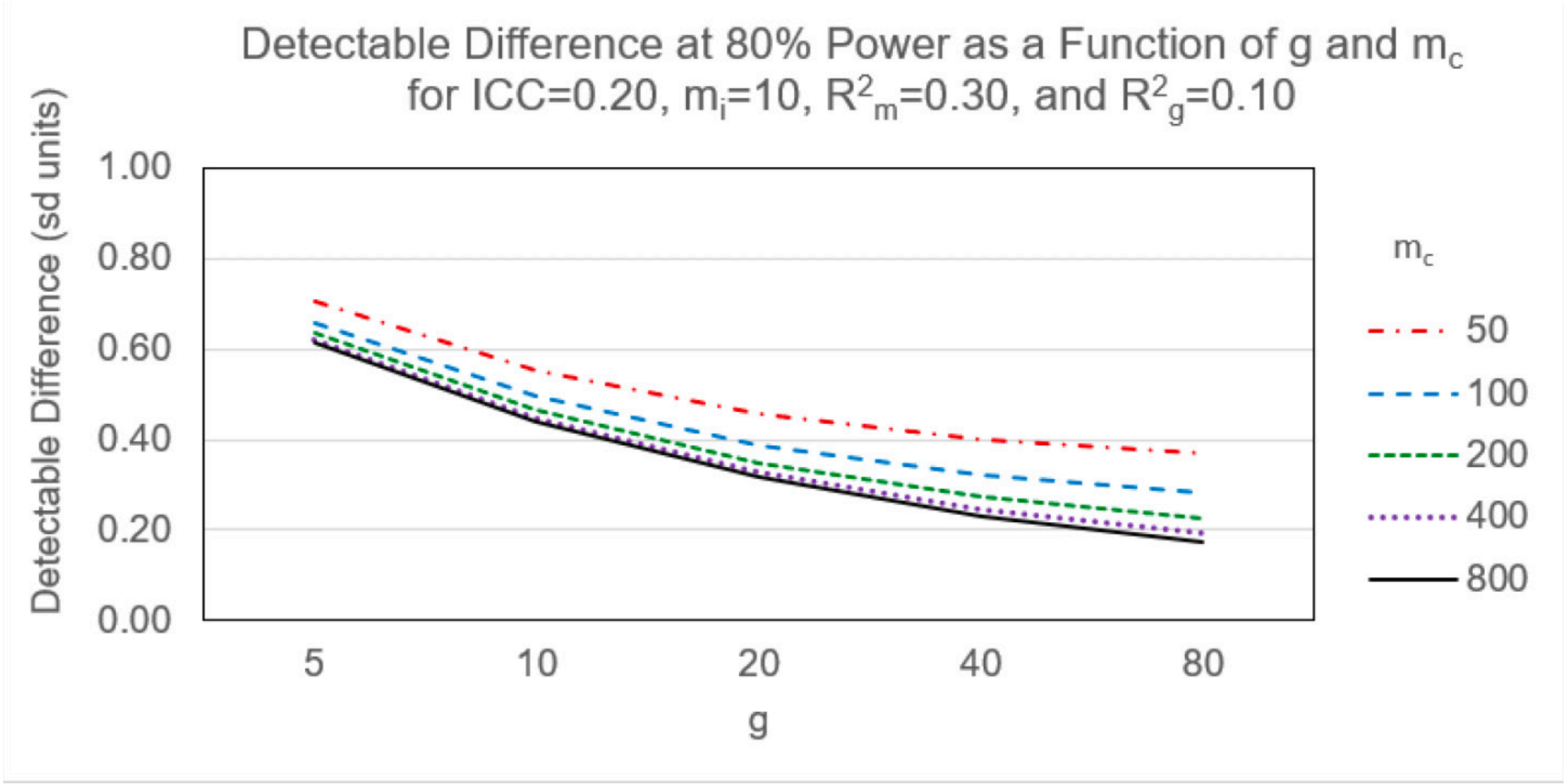
Continuous: **variance = 1.00**

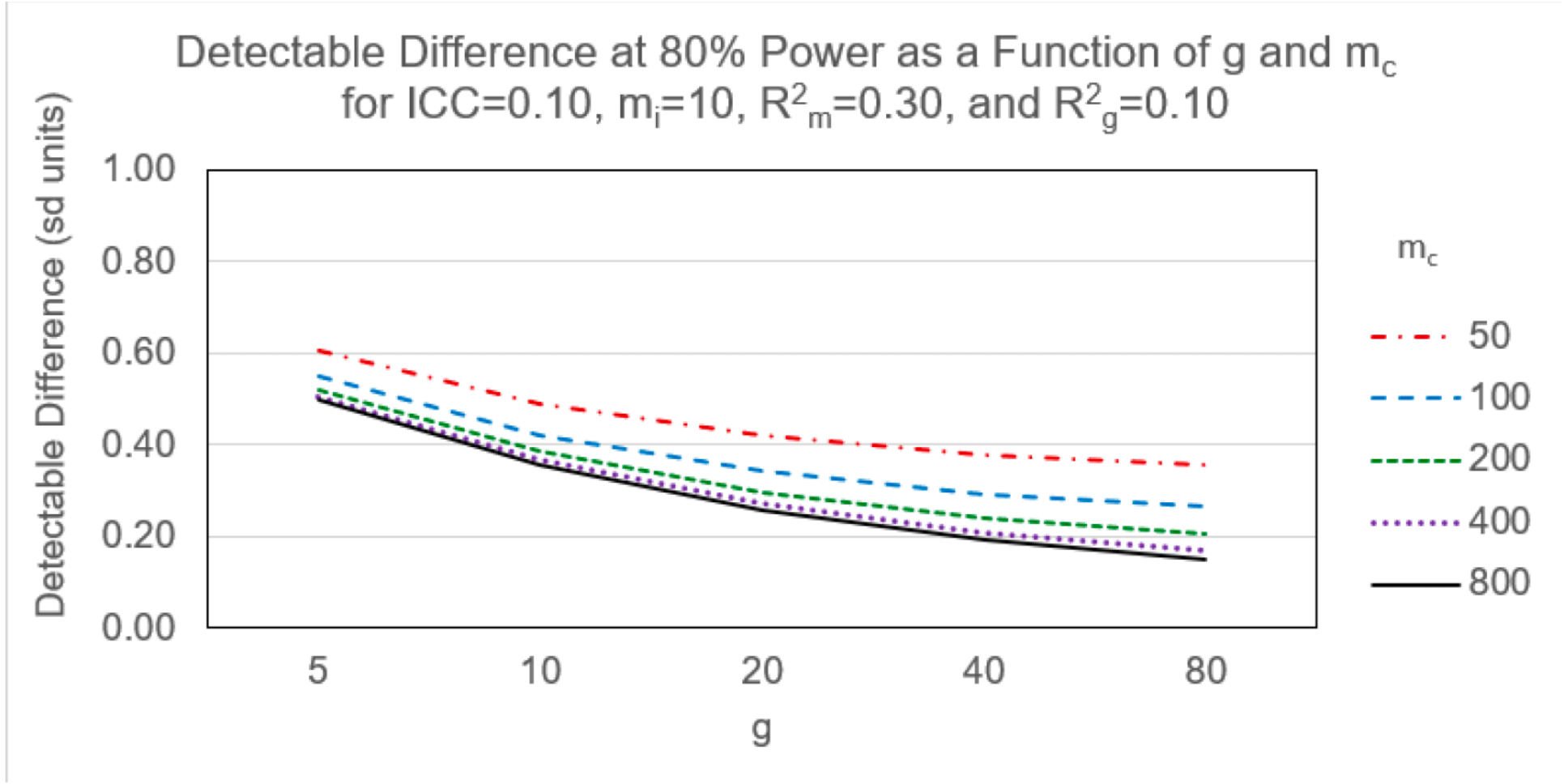
Design and Analytic Plan

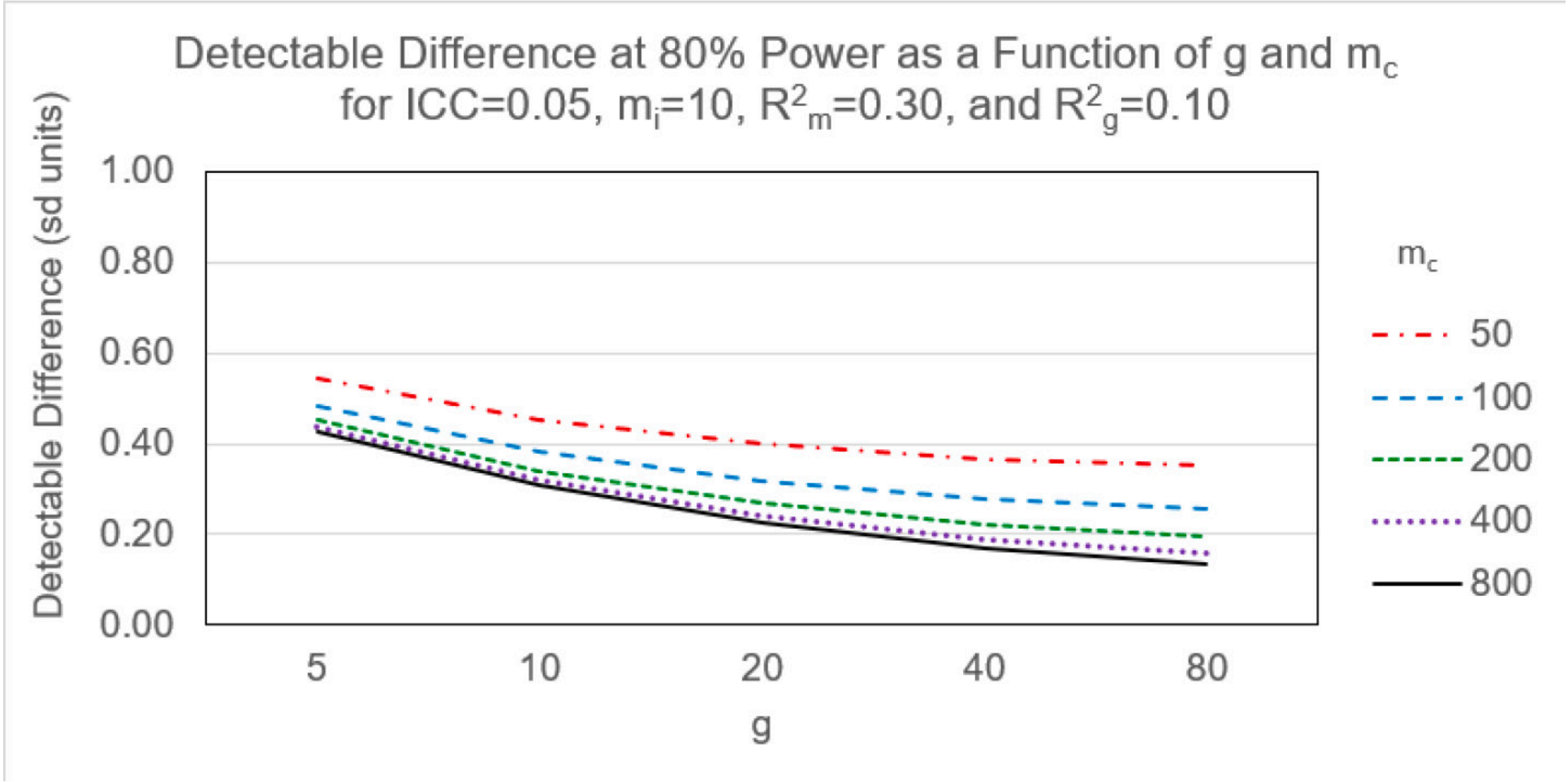
Analysis of a simple difference

Analysis of a Simple Difference

Absolute Detectable Difference					
Members in the Control Condition	**Groups in the Intervention Condition**				
	3	**4**	**5**	**8**	**10**
13	0.9490	0.8883	0.8490	0.7844	0.7609
25	0.7502	0.6931	0.6563	0.5962	0.5744
50	0.6475	0.5855	**0.5449**	0.4774	0.4526
100	0.5948	0.5285	0.4844	0.4093	0.3810
200	0.5678	0.4987	0.4522	0.3718	0.3407


Detectable Difference at 80% Power as a Function of g and m_c
for ICC=0.20, m_i=10, R^2_m=0.30, and R^2_g=0.10
Detectable Difference (sd units)
1.00
0.80
0.60
0.40
0.20
0.00
5
10
20
40
80
g
m_c
50
100
200
400
800


Detectable Difference at 80% Power as a Function of g and m_c
for ICC=0.10, m_i=10, R^2_m=0.30, and R^2_g=0.10
Detectable Difference (sd units)
1.00
0.80
0.60
0.40
0.20
0.00
5
10
20
40
80
g
m_c
50
100
200
400
800


Detectable Difference at 80% Power as a Function of g and m_c
for ICC=0.05, m_i=10, R^2_m=0.30, and R^2_g=0.10
Detectable Difference (sd units)
1.00
0.80
0.60
0.40
0.20
0.00
5
10
20
40
80
g
m_c
50
100
200
400
800

Summary

- A parallel GRT remains the best comparative design available whenever the investigator wants to evaluate an intervention that…
  - operates at a group level
  - manipulates the social or physical environment
  - cannot be delivered to individuals
- IRGTs are widely used and rarely recognized as such, with most designed and analyzed ignoring the ICC created through the method of intervention delivery.
- Investigators new to GRTs, and IRGTs should collaborate with more experienced colleagues, especially experienced methodologists.
- Research Methods Resources Website
  - https://researchmethodsresources.nih.gov/
  - Material and sample size calculators for parallel GRTs and IRGTs.

Coming Later This Year

- We will add a section on Stepped Wedge Group- or Cluster-Randomized Trials to the Research Methods Resources website later this year.
- The section will be structured like the GRT and IRGT sections with background, references, frequently asked questions, and a sample size calculator.