Keynote Address

Pragmatic Trials involving Group Randomization or Delivery of Interventions to Groups: GRTs, IRGTs, and SW-GRTs

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

Design & Analysis of Embedded Pragmatic Clinical Trials
May 2, 2019

Pragmatic Trials

- Pragmatic and explanatory trials were first described by Schwartz & Lellouch (1967).
  - Explanatory trials test causal research hypotheses.
  - Pragmatic trials help users choose between options for care.
- Similar to efficacy and effectiveness trials (Cochrane, 1971).
  - Efficacy trials evaluate an intervention under controlled conditions.
  - Effectiveness trials evaluate an intervention under real-world conditions.

- Schwartz, D., & Lellouch, J. Explanatory and pragmatic attitudes in therapeutical trials. Journal of Chronic Diseases, 1967, 20(8), 637-648. PMID4860352
- Cochrane, A.L. Effectiveness and efficacy: random reflections on health services. Nuffield Provincial Hospitals Trust, London, 1971. (cited in Flay, Brian R. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. Preventive Medicine, 1986, 15(5), 451-474. PMID3534875.

Pragmatic Trials and Electronic Health Records

- EHRs provide the opportunity to collect data for research purposes at a significantly reduced cost compared to traditional methods.
- However…
  - The data are not collected for research purposes.
  - The EHR may not have all the variables needed for the research.
  - The investigators may have limited control over the EHR.
- But if they have some influence…
  - EHRs provide an efficient way to target populations of interest.
  - They can facilitate recruitment.
  - Participation can be flagged and the EHR can be used to facilitate delivery of the intervention.
  - EHR can be used to passively capture a variety of data.

Methodological Considerations

- Pragmatic trials do not necessarily require a different set of research designs, measures, analytic methods, etc.
- As always, the choice of methods depends on the research question.
- The research question dictates
  - the intervention, target population, and variables of interest,
  - which dictate the setting, research design, measures, and analytic methods.
- Randomized trials will provide the strongest evidence.
  - Which kind of randomized trial will depend on the research question.
- Alternatives to randomized trials are also available.

Three Kinds of Randomized Trials

- Randomized Clinical Trials (RCTs)
  - Individuals randomized to study conditions with no interaction among participants after randomization and no 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.

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.

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.

Impact on the Design

- 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.
- Stepped wedge GRTs
  - Crossing of groups with study conditions avoids most confounding.
  - However, intervention effects are confounded with calendar time, as more groups are in the intervention condition as the study progresses.
  - SW-GRTs are inherently less rigorous than parallel GRTs and should be considered only when a parallel GRT is not appropriate.

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.

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

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

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

Impact on the Analysis in a GRT or IRGT

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

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

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

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

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

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

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,

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

  - When group membership is not established by random assignment,

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

  - Or equivalently,

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

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.

Impact on the Analysis for SW-GRTs

- Crossing of groups with study conditions often reduces the impact of the ICC compared to a parallel GRT, either improving power or allowing a smaller study.
- There are other potential sources of bias in the SW-GRT:
  - The intervention is confounded with time.
  - The intervention effect may vary over time.
  - The intervention effect may vary by group.
  - Patterns of correlation may vary over time.
- Any analysis that assumes that the intervention effect is constant over time and across groups, and that the pattern of correlation is constant, may be biased.
- Compared to a parallel GRT, SW-GRTs are at greater risk to the effects of external events that affect the outcomes of the trial.

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)

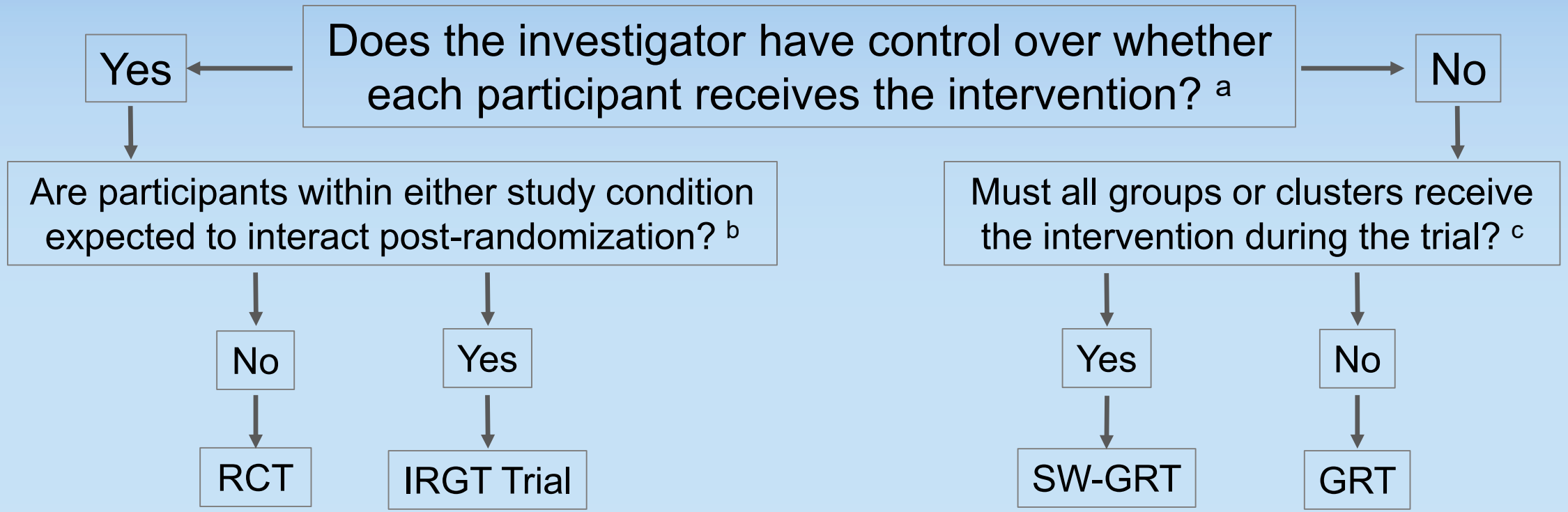
- Though Cornfield's remarks were addressed only to GRTs, they also apply to IRGTs, and to SW-GRTs

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

The Need for GRTs, IRGTs, and SW-GRTs

- 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
- An SW-GRT is an alternative to a parallel GRT if…
  - Preliminary evidence makes it unethical to withhold the intervention.
  - It is impossible to implement the intervention in all groups simultaneously.
  - External events are unlikely to affect the outcomes before the end of the trial.

Choosing Among These Design Options



[a] Control is lacking if the intervention cannot be delivered to individuals or if the risk of contamination is high.

[b] Interaction could occur if the intervention is delivered through a physical group, a virtual group, or though interventionists who work with multiple participants; such interaction can occur in either or both conditions.

[c] The may be ethical, political, or logistical reasons to deliver the intervention to all groups before the end of the trial.

Preferred Analytic Models for Standard 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 Standard 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).
  - 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.

What About Individually Randomized Group Treatment Trials (IRGTs)?

- 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.
    - Base df on the number of groups, not the number of members.
  - Mixed models are the most common approach.

- 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. Psychl Methods. 2011;16(2):149-65. PMID21517179.

What About GRTs or IRGTs In Which Members Belong to More than one Group or Change 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 multilevel, 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 in 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.

What About Stepped Wedge Designs (SW-GRTs)?

- The original Hussey & Hughes (2007) approach assumed a common secular trend and an immediate and constant intervention effect.
- Hughes et al. (2015) allow the treatment effects to vary across groups or over time.
- Hooper et al. (2016) allow the between-period ICC to be less than the within-period ICC, but allow no further decay.

- Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clinical Trials. 2007;28(2):182-91. PMID16829207.
- Hughes JP, Granston TS, et al. Current issues in the design and analysis of stepped wedge trials. Contemp Clinical Trials. 2015;45(Pt A):55-60. PMID26247569.
- Hooper R, Teerenstra S, et al. Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. Stat Med. 2016;35(26):4718-28. PMID27350420.

What About Stepped Wedge Designs (SW-GRTs)?

- Kasza et al. (2017) allow the between-period ICC to decay steadily.
- Grantham et al. (2019) allow even more flexible decay models.

- Kasza J, Hemming K et al. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. Stat Meth in Med Res. 2017;0(0)1-14. PMID29027505.
- Grantham KL, Kasza J, et al. Accounting for a decaying correlation structure in cluster randomized trials with continuous recruitment. Stat Med. 2019;38(11):1918-34. PMID30663132.

State of the Science for Analytic Methods in GRTs, IRGTs, and SW-GRTs

- GRTs, IRGTs, and SW-GRTs require analyses that reflect their nested or partially-nested and sometimes cross-classified designs.
- Used alone, the usual methods based on the General or Generalized Linear Model are not valid.
- Methods based on the General Linear Mixed Model and on the Generalized Linear Mixed Model are widely applicable.
- Other methods can be used effectively, with proper care, including randomization tests, GEE, and two-stage methods.

Power for Group-Randomized Trials

- The usual methods must be adapted to reflect the nested design
  - The variance is greater in a 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 GRT.
- Power in 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

- Methods for sample size estimation for IRGTs have been published.
  - Pals SP, 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 in Med. 2013;32(1):81-98. PMID22865729.
  - Moerbeek M, Teerenstra S. Power analysis of trials with multilevel data. Boca Raton: CRC Press; 2016.

Power for SW-GRTs

- Methods for sample size estimation for SW-GRTs have been published.
  - Moerbeek M, Teerenstra S. Power analysis of trials with multilevel data. Boca Raton: CRC Press; 2016.
  - Hemming K, Taljaard M. Sample size calculations for stepped wedge and cluster randomised trials: a unified approach. J Clin Epi. 2016;69:137-46. PMID26344808.
  - Hooper R, Teerenstra S, et al. Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. Stat in Med. 2016;35(26):4718-28. PMID27350420.
  - Kasza J, Hemming K et al. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. Stat Meth in Med Res. 2017;0(0)1-14. PMID29027505.
  - Li F, Turner EL, et al. Sample size determination for GEE analyses of stepped wedge cluster randomized trials. Biometrics. 2018;74(4):1450-8. PMID29921006.

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
    - 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 on GRTs and IRGTs and a sample size calculator for GRTs.