Group-Randomized Trials in Public Health and Medicine

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

SCT Webinar
August 16, 2016

Disclosures

- David M. Murray, Ph.D.
- Group-Randomized Trials In Public Health and Medicine
- No commercial interests
- No discussion of off-label drug usage
- No relevant financial relationships

Three Kinds of Randomized Trials

- Individually Randomized Clinical Trials (RCTs)
  - Individuals randomized to study conditions with no connection among participants after randomization
    - Most surgical and drug trials, some behavioral trials
- Individually Randomized Group Treatment Trials (IRGTs)
  - Individuals randomized to study conditions with some connection among participants after randomization
    - Many behavioral trials
- Group-Randomized Trials (GRTs)
  - Groups randomized to study conditions with some connection among participants before and after randomization
    - Many trials conducted in communities, worksites, schools, etc.

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 most IRGTs have N<100. Confounding can be more of a concern in IRGTs than in RCTs.
- Group-randomized trials
  - GRTs often involve a limited number of groups, often <50.
  - There may be limited opportunity for randomization to distribute potential confounders evenly.
  - Confounding is usually a concern in GRTs if G is <50.

Impact on the Analysis

- 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 created for an IRGT will develop similar connections.
- Those connections will create a positive intraclass correlation that reflects extra variation attributable to the group.

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

Impact on the Analysis

- 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 and Power

- The variance of any group-level statistic will be larger.
- The df to estimate the ICC 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.
- Nested factors must be modeled as random effects (Zucker, 1990), including groups and facilitators, if nested.

- Zucker DM. An analysis of variance pitfall: The fixed effects analysis in a nested design. Educational and Psychological 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. American Journal of Epidemiology. 1978;108(2):100-2.

The Need for GRTs and IRGTs

- A GRT remains the best comparative design available when the investigator wants to evaluate an intervention that…
  - operates at a group level
  - manipulates the social or physical environment
  - cannot be delivered to individuals without contamination
- An IRGT is the best comparative design when...
  - Individual randomization is possible without contamination
  - There are good reasons to deliver the intervention in groups
- The challenge is to create trials that are:
  - Rigorous enough to avoid threats to validity of the design,
  - Analyzed so as to avoid threats to statistical validity,
  - Powerful enough to provide an answer to the question,
  - And inexpensive enough to be practical.

Important References on GRTs, IRGTs

- 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.
- Hayes RJ, Moulton LH. Cluster Randomised Trials. Boca Raton, FL: Taylor & Francis Group, LLC; 2009.
- Campbell MJ, Walters SJ. How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research. Chichester: John Wiley & Sons Ltd.; 2014.
- Pals SP, Murray DM et al. Individually randomized group treatment trials: a critical appraisal of frequently used design and analytic approaches. American Journal of Public Health. 2008;98(8):1418-24.
- Baldwin SA, Bauer DJ, et al. Evaluating models for partially clustered designs. Psychological Methods. 2011;16(2):149-65.
- Roberts C, Walwyn R. Design and analysis of non-pharmacological treatment trials with multiple therapists per patient. Statistics in Medicine. 2013;32(1):81-98.

Planning the Trial

- The driving force must be the research question.
  - The question will identify the target population, the setting, the endpoints, and the intervention.
  - Those factors will shape the design and analysis plan.
- The primary criteria for choosing that question should be:
  - Is it important enough to do?
    - Will the trial address an important public health question?
    - Will the results advance the field?
  - Is this the right time to do it?
    - Is there preliminary evidence of feasibility and efficacy for the intervention?
    - Are there good estimates for the parameters needed to size the study?
- The investigators should keep the question in mind.

Strategies to Limit Threats to Internal Validity

- Randomization
- A priori matching or stratification
  - Of groups in GRTs, of members in IRGTs and RCTs
- Objective measures
- Independent evaluation personnel who are blind to conditions
- Analytic strategies
  - Regression adjustment for covariates
- Avoid the pitfalls that invite threats to internal validity
  - Testing and differential testing
  - Instrumentation and differential instrumentation
  - Regression to the mean and differential regression to the mean
  - Attrition and differential attrition

Strategies to Protect the Validity of the Analysis

- Avoid model misspecification
  - Plan the analysis concurrent with the design.
  - Plan the analysis around the primary endpoints.
  - Anticipate all sources of random variation.
  - Anticipate patterns of over-time correlation.
  - Consider alternate models for time.
  - Assess potential confounding and effect modification.

Strategies to Protect the Validity of the Analysis

- Avoid low power
  - Employ strong interventions with good reach.
  - Maintain reliability of intervention implementation.
  - Employ more and smaller groups instead of a few large groups.
  - Employ more and smaller surveys or continuous surveillance instead of a few large surveys.
  - Employ regression adjustment for covariates to reduce variance and intraclass correlation.

Factors That Can Reduce Precision

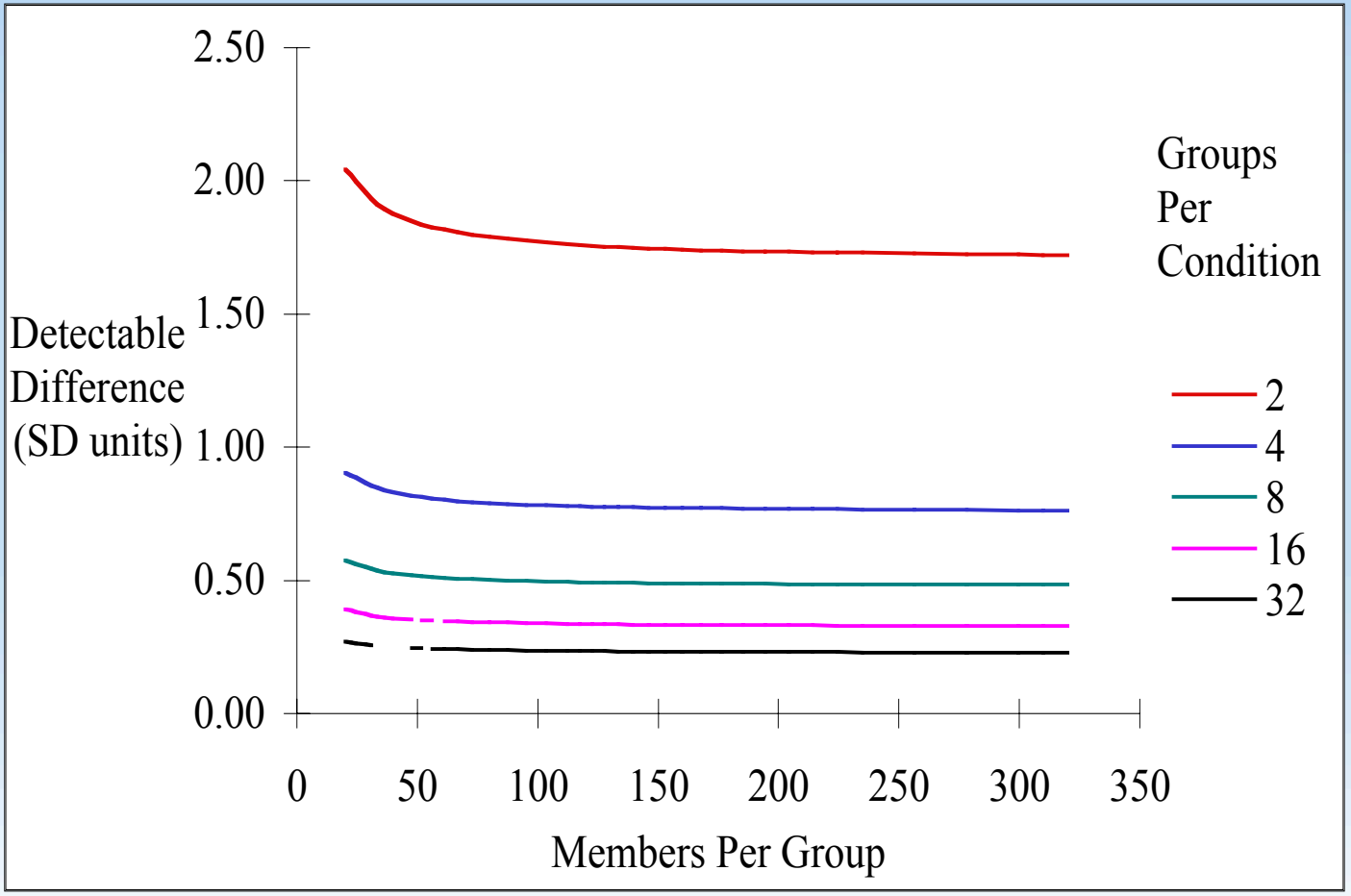
- The variance of the condition mean in a GRT is:

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

- 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

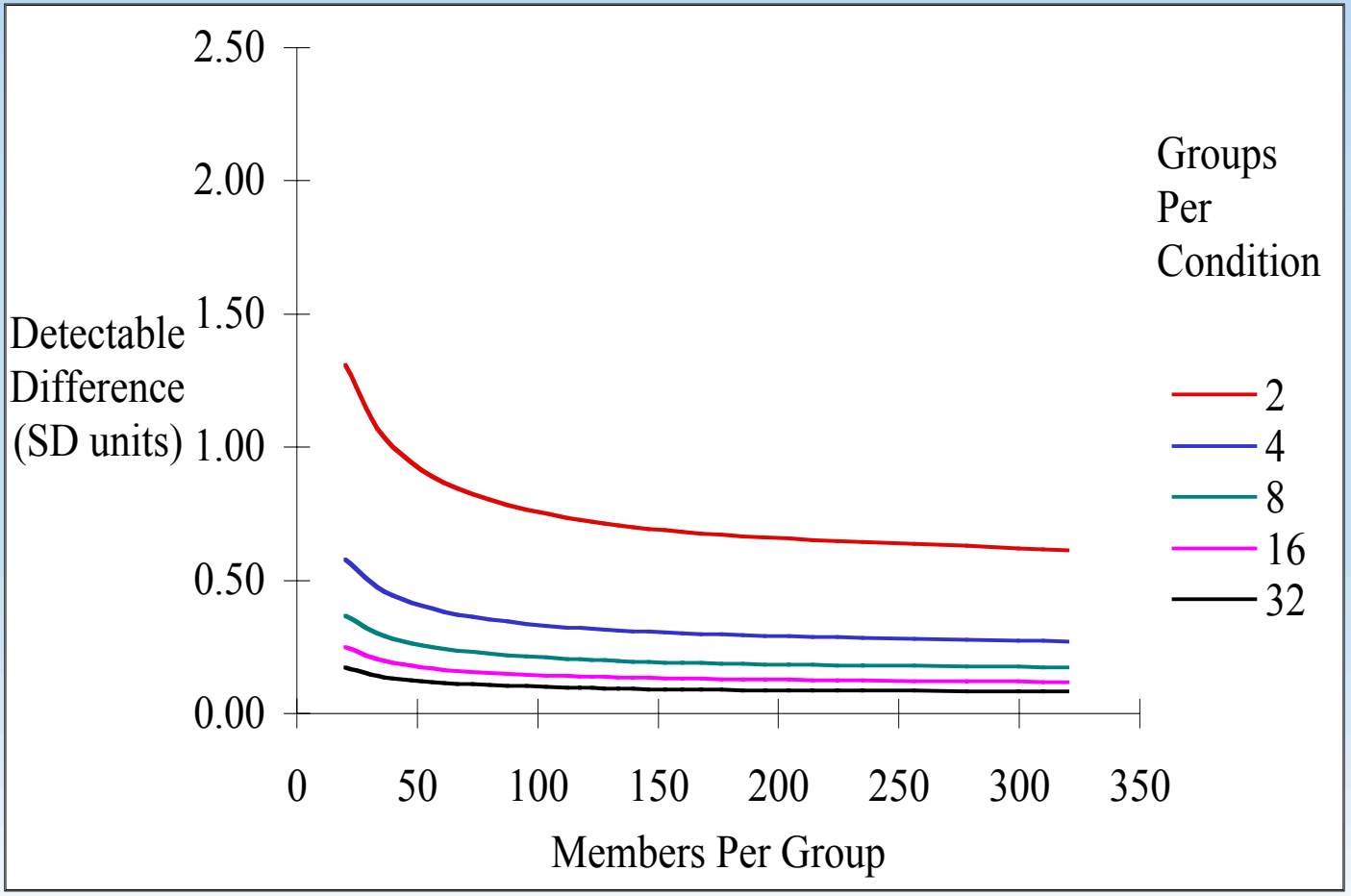
Strategies to Improve Precision

- Increased replication (ICC=0.100)



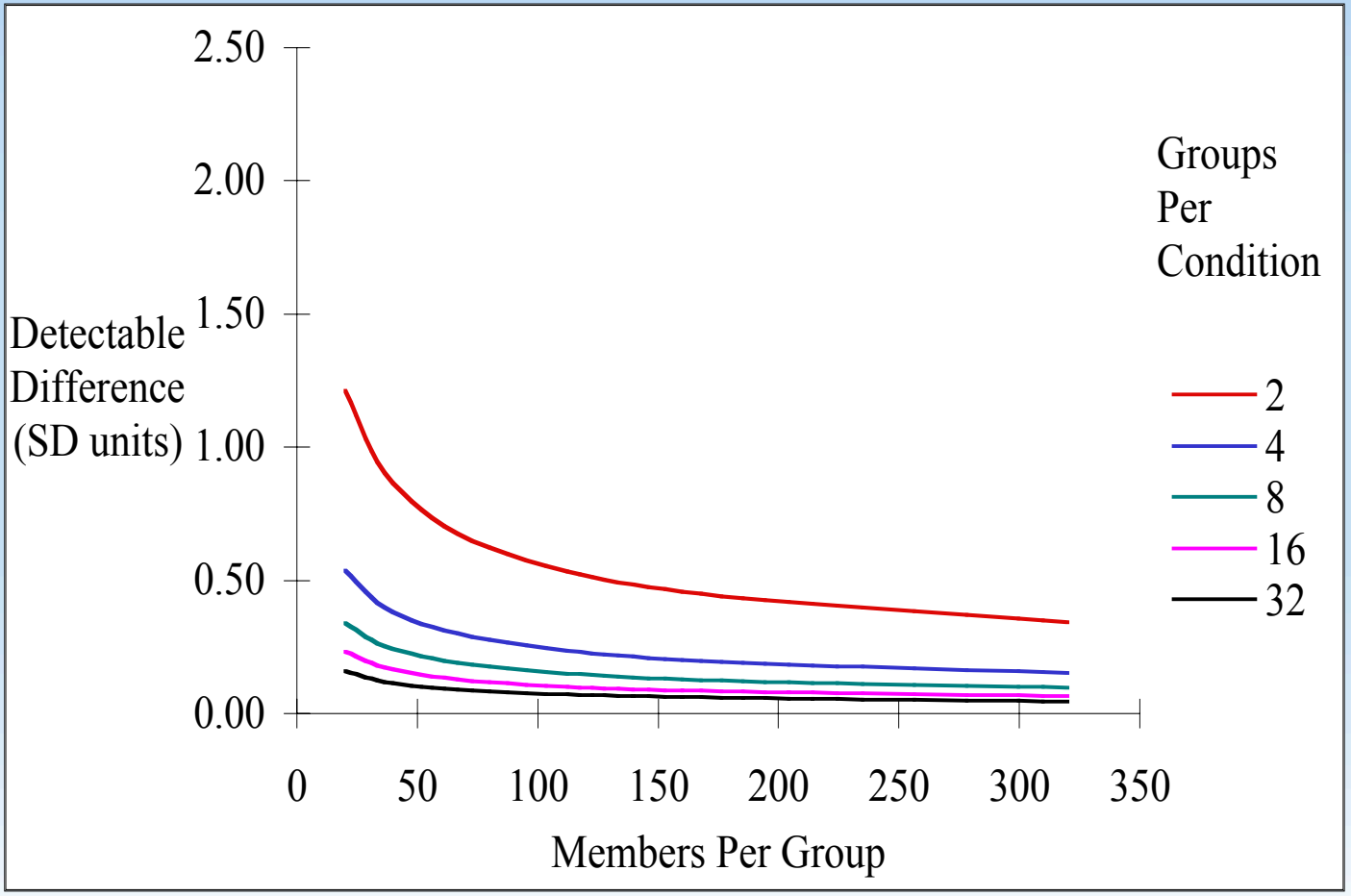
Strategies to Improve Precision

- Reduced ICC (ICC=0.010)



Strategies to Improve Precision

- The law of diminishing returns (ICC=0.001)



Single Factor and Factorial Designs

- Most GRTs involve only one treatment factor.
  - Condition
- Most GRTs have only two levels of that treatment factor.
  - Intervention vs. control
- Most GRTs cross Condition with Time.
  - Nested cohort designs
  - Nested cross-sectional designs
- Some GRTs include stratification factors.
  - Multi-center GRTs cross Condition with Field Center.
  - Single-center GRTs often stratify on factors related to the outcome or to the ease of implementation of the intervention.

Time as a Factor

- Posttest-only design
- Pretest-posttest design
- Extended designs
  - Additional discrete time intervals before and/or after intervention
  - Continuous surveillance

Cross-Sectional and Cohort Designs

- Nested cohort design
  - The research question involves change in specific members.
  - Measure the same members at each time data are collected.
- Nested cross-sectional design
  - The research question involves change in an entire population.
  - Select a new sample of members each time data are collected.

Cross-Sectional and Cohort Designs

- Strengths and weaknesses

Cross-section	*Cohort*
in migration and out migration	mortality
group change	individual change
recruitment costs	tracking and follow-up costs
less powerful?	more powerful?
full dose?	full dose?

A Priori Matching or Stratification

- Rationale
  - Either can be used if the investigators want to ensure balance on an important potential source of bias.
  - A priori stratification is preferred if the investigators expect the intervention effect to be different across strata.
  - A priori matching is useful if the matching factors are well correlated with the primary endpoint.
  - The choice of matching vs. stratification will often depend on the number of groups available and on the expected correlation.
  - Work by Donner et al. (2007) favors stratification when m<100.

- Donner A, Taljaard M, Klar N. The merits of breaking the matches: a cautionary tale. Statistics in Medicine. 2007;26(9):2036-51.

Planning the Trial
Summary

- All the design features common to RCTs are available to GRTs and IRGTs, with the added complication of an extra level of nesting:
  - Nested cohort and nested cross-sectional designs;
  - Post only, pre-post, and extended designs;
  - Single factor designs and factorial designs;
  - A priori matching or stratification, and post hoc stratification;
  - Constrained randomization
- The primary threats to internal and statistical validity are well known, and defenses are available.
  - Plan the study to reflect the nested design, with sufficient power for a valid analysis, and avoid threats to internal validity.

A Classification Scheme for Statistical Models

	Gaussian Distribution	Non-Gaussian Distribution
One Random Effect	General Linear Model	Generalized Linear Model
Two Or More Random Effects	General Linear Mixed Model	Generalized Linear Mixed Model

- Fixed effect: the investigators want to draw inferences only about the levels used in the study.
- Random effect: the investigators want to draw inferences about some larger population of levels that are only represented by the levels used in the study.

Preferred Analytic Strategies for 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.

Preferred Analytic Strategies for 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.
  - 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.

What About Randomization Tests?

- The intervention effect is a function of unadjusted or adjusted group-specific means, slopes or other group-level statistic.
- Under the null hypothesis of no intervention effect, the actual arrangement of those group-level statistics among the study conditions is but one of many equally likely arrangements.
- The randomization test systematically computes the effect for all possible arrangements.
- The probability of getting a result more extreme than that observed is the proportion of effects that are greater than that observed.
- No distributional or other assumptions are required.

What About Randomization Tests?

- Strengths
  - Gail et al. (1996) found that randomization tests had nominal Type I and II error rates across conditions common to GRTs.
    - Even when the member-level errors were non-normal,
    - Even when very few heterogeneous groups are assigned to each condition,
    - Even when the ICC was large or small,
    - So long as there was balance at the level of the group.
  - Programs for randomization tests are available in print and on the web.

- Gail MH, Mark SD, Carroll RJ, Green SB, Pee D. On design considerations and randomization-based inference for community intervention trials. Statistics in Medicine. 1996;15(11):1069-92.

What About Randomization Tests?

- Weaknesses
  - The unadjusted randomization test does not offer any more protection against confounding than other unadjusted tests (Murray et al., 2006).
  - Randomization tests provide only a point estimate and a p-value.
  - Regression adjustment for covariates requires many of the same assumptions as the model-based tests.

- Murray DM, Hannan PJ, Varnell SP, McCowen RG, Baker WL, Blitstein JL. A comparison of permutation and mixed-model regression methods for the analysis of simulated data in the context of a group-randomized trial. Statistics in Medicine. 2006;25(3):375-88.

What About Randomization Tests?

- Model-based methods provide parameter estimates, standard errors, and the nominal Type I error rate (Murray et al., 2006).
  - Even if the member- or group-level errors were non-normal, unless they were very skewed or heavy tailed (unpublished dissertation).
  - Even when few heterogeneous groups were assigned to each condition.
  - Even when the ICC was large or small.
  - So long as there was balance at the level of the group.
  - Randomization tests and model-based tests perform similarly under most conditions.
  - Randomization tests are preferred for very skewed or heavy tailed distributions.

What About a Method Like GEE That is Robust Against Misspecification?

- Methods based on GEE use an empirical sandwich estimator for standard errors.
- That estimator is asymptotically robust against misspecification of the random-effects covariance matrix.
- When the degrees of freedom are limited (<40), the empirical sandwich estimator has a downward bias.
- Recent work provides corrections for that problem; several have recently been incorporated into SAS PROC GLIMMIX (beginning with SAS 9.1.3).

- Methods that employ the corrected empirical sandwich estimator may have broad application in GRTs.

What About Fixed-Effect Methods in Two Stages?

- Introduced as the a solution for nested designs in the 1950s.
  - Commonly known as the means analysis.
  - Simple to do and easy to explain.
  - Gives results identical to the mixed-model ANOVA/ANCOVA if both are properly implemented.
  - Can be adapted to perform random coefficients analyses.
  - Can be adapted to complex designs where one-stage analyses are not possible.
  - Used in several large trials, including CATCH, MHHP, REACT, CYDS, and TAAG.
- Two-staged models can be very useful in GRTs.

What About Deleting the Unit of Assignment From the Model if it is not Significant?

- The df for such tests are usually limited; as such, their power is usually limited.
- Standard errors for variance components are not well estimated when the variance components are near zero.
- Even a small ICC, if ignored, can inflate the Type I error rate if the number of members per group is moderate to large.

- The prudent course is to retain all random effects associated with the study design and sampling plan.

What About Unbalanced Designs?

- Group-level imbalance can create analytic problems (Gail et al., 1996; Murray et al., 2006).
- Member-level imbalance can create Type I error inflation and the risk increases with the level of imbalance.
- Johnson et al. (2015) compared 10 model-based approaches to member imbalance.
  - A one-stage mixed model with Kenward-Roger df and unconstrained variance components performed well for g≥14.
  - A two-stage model weighted by the inverse of the estimated theoretical variance of the group means and with unconstrained variance components performed well for g≥6.

- Johnson JL, Kreidler SM, Catellier DJ, Murray DM, Muller KE, Glueck DH. Recommendations for choosing an analysis method that controls Type I error for unbalanced cluster sample designs with Gaussian outcomes. Statistics in Medicine. 2015;34(27): 3531-45.

What About Constrained Randomization?

- Li et al. (2015) evaluated model-based and randomization tests in the context of constrained randomization in a GRT.
  - The unadjusted randomization test maintained the nominal Type I error rate; the unadjusted model-based test was conservative.
  - Adjusted model-based and randomization tests were similar.
    - Both maintained the nominal Type I error rate.
    - Both had better power under constrained randomization.
  - Correct specification of the permutation distribution is essential under constrained randomization.
- Constrained randomization can improve power if used well.

- Li F, Lokhnygina Y, Murray DM, Heagerty PJ, DeLong ER. An evaluation of constrained randomization for the design and analysis of group-randomized trials. Statistics in Medicine. 2015;35(10):1565-79. PMC4826850.

Is the Non-Negativity Constraint OK?

- Software based on maximum likelihood routinely constrains variance estimates to be non-negative.
  - Combined with traditional methods for calculating df, this constraint introduces a positive bias in the variance component estimates and depresses the Type I error rate, often dramatically (Swallow & Monahan, 1984; Murray et al., 1996).
  - Earlier advice was to avoid the non-negativity constraint.
- Recent evidence suggests that the Kenward-Roger method for df addresses this problem (Andridge et al., 2014).

- Swallow WH, Monahan JF. Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components. Technometrics. 1984;26(1):47-57.
- Andridge RR, Shoben AB, Muller KE, Murray DM. Analytic methods for individually randomized group treatment trials and group-randomized trials when subjects belong to multiple groups. Statistics in Medicine. 2014;33(13):2178-90. PMC4013262.

What About Individually Randomized Group Treatment Trials (IRGTs)?

- Many studies randomize participants as individuals but deliver treatments in small groups (cf. Pals et al., 2008).
  - Psychotherapy, weight loss, smoking cessation, etc.
  - Participants nested within groups, facilitators nested within conditions
  - Little or no group-level ICC at baseline.
  - Positive ICC later, with the magnitude proportional to the intensity and duration of the interaction among the group members.

- Pals SP, Murray DM, Alfano CM, Shadish WR, Hannan PJ, Baker WL. Individually randomized group treatment trials: a critical appraisal of frequently used design and analytic approaches. American Journal of Public Health. 2008;98(8):1418-24. PMC2446464
- Pals SL, Murray DM, Alfano CM, Shadish WR, Hannan PJ, Baker WL. Erratum. American Journal of Public Health. 2008;98(12):2120.

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 small groups.
    - Base df on the number of small groups, not the number of members.

- Baldwin SA, Bauer DJ, Stice E, Rohde P. Evaluating models for partially clustered designs. Psychological Methods. 2011;16(2):149-65. PMC3987820.

What About IRGTs In Which Members Belong to More than one Group or Change Groups?

- The IRGT literature assumes that each member belongs to a single group and that group membership does not change.
  - That pattern is not likely to hold in practice.
  - Andridge (2014) found that failure to account for multiple group membership can inflate Type I error for the methods described thus far.
  - Roberts (2013) found that multiple membership multilevel models address this problem.
    - They require data on membership time in each group, which is not routinely collected in IRGTs.

- Andridge RR, Shoben AB, Muller KE, Murray DM. Analytic methods for individually randomized group treatment trials and group-randomized trials when subjects belong to multiple groups. Statistics in Medicine. 2014;33(13):2178-90. PMC4013262.
- Roberts C, Walwyn R. Design and analysis of non-pharmacological treatment trials with multiple therapists per patient. Statistics in Medicine. 2013;32(1):81-98.

Analytic Strategies
Summary

- GRTs require analyses that reflect the nested designs inherent in these studies.
- 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.
  - For designs having one or two time intervals, mixed-model ANOVA/ANCOVA is recommended.
  - For designs having three or more time intervals, random coefficients models are recommended.
- Other methods can be used effectively, with proper care, including randomization tests, GEE, and two-stage methods.

Analytic Strategies
Summary

- Other approaches are not appropriate, including analysis at a subgroup level, deleting the unit of assignment if it or the ICC is not significant, designs with one group per condition, and Kish's effective df.
- Unbalanced designs can create analytic problems and an inflated Type I error rate; special methods are required.
- Constrained randomization can be helpful.
- IRGTs face similar problems to GRTs and the solutions are similar: model the small groups or common change agents as nested random effects, with implications for df and testing.

Power for Group-Randomized Trials

- The usual methods must be adapted for the nested design
  - A good source on power is Chapter 9 in Murray (1998).
  - Other texts include Donner & Klar, 2000; Hayes & Moulton, 2009; Campbell & Walters, 2014.
  - Recent review articles include Gao et al. (2015) and Rutterford et al. (2015).

- 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.
- Hayes RJ, Moulton LH. Cluster Randomised Trials. Boca Raton, FL: Taylor & Francis Group, LLC; 2009.
- Campbell MJ, Walters SJ. How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research. Chichester: John Wiley & Sons Ltd.; 2014.
- Gao F, Earnest A, Matchar DB, Campbell MJ, Machin D. Sample size calculations for the design of cluster randomized trials: A summary of methodology. Contemporary Clinical Trials. 2015;42:41-50.
- Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. International Journal of Epidemiology. 2015;44(3):1051-67. PMC4521133.

Power for Individually Randomized Group-Treatment Trials

- Power for IRGTs is often even trickier, and the literature is more limited (cf. Pals et al. 2008; Heo et al., 2014)

- Pals SP, Murray DM, Alfano CM, Shadish WR, Hannan PJ, Baker WL. Individually randomized group treatment trials: a critical appraisal of frequently used design and analytic approaches. American Journal of Public Health. 2008;98(8):1418-24. PMC2446464
- Pals SL, Murray DM, Alfano CM, Shadish WR, Hannan PJ, Baker WL. Erratum. American Journal of Public Health. 2008;98(12):2120.
- Roberts C, Walwyn R. Design and analysis of non-pharmacological treatment trials with multiple therapists per patient. Statistics in Medicine. 2013;32(1):81-98.
- Heo M, Litwin AH, Blackstock O, Kim N, Arnsten JH. Sample size determinations for group-based randomized clinical trials with different levels of data hierarchy between experimental and control arms. Statistical Methods in Medical Research. 2014. PMC4329103.

Cornfield's Two Penalties

- Extra variation
  - Condition-level statistic vs. group-level statistic
  - Greater variation in the group-level statistic
  - Reduced power, other factors constant.
- Limited df
  - df based on the number of groups
  - Number of groups in a GRT is often limited
  - Reduced power, other factors constant

- Cornfield J. Randomization by group: a formal analysis. American Journal of Epidemiology. 1978;108(2):100-2.

Unbalanced Designs

- As long as the ratio of the largest to the smallest group is no worse than about 2:1, the standard methods are fine.
- Given more extreme imbalance, other methods are required.
  - For a GRT, several recent papers provide alternative methods.
    - van Breukelen G, Candel M, Berger M. Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. Statistics in Medicine. 2007;26(13):2589-603.
    - 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. Statistics in Medicine. 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. Statistics in Medicine. 2016;35(12):2000-15.
  - For an IRGT, see
    - 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. Statistics in Medicine. 2009;28(18):2307-24.

Power
Summary

- The usual methods for detectable difference, sample size, and power must be adapted to reflect the nested design.
- Power for GRTs and IRGTs is tricky.
- Both of Cornfield's penalties must be addressed: extra variation and limited df.
- Failure to do so will result in an inflated Type I error.
- The most important factors affecting power in a GRT are the ICC and the number of groups per condition.
- Investigators should seek good estimates for these parameters, drawn from circumstances as close to the trial under consideration as possible.

Power
Summary

- Those factors are also important in IRGTs, but there are additional factors as well
  - Whether there are small groups in more than one arm,
  - Whether members belong to more than one small group at the same time,
  - Whether group membership changes over time,
  - Whether leaders are crossed or nested within conditions,
  - Whether the small group interaction is limited or intense.
- It is even more important in an IRGT for the investigator to seek collaboration with a good methodologist.

Summary

- A 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
- GRTs provide better quality evidence and are either more efficient or take less time than the alternatives.
- Even so, GRTs are more challenging than the usual RCT.
- IRGTs present many of the same issues found in GRTs.
- Investigators new to GRTs and IRGTs should collaborate with more experienced colleagues, especially experienced methodologists.