Design and Analysis of Individually Randomized Group Treatment (IRGT) Trials

Sherri L. Pals, Ph.D.

Mathematical Statistician
Center for Global Health, Division of HIV/AIDS & TB
U.S. Centers for Disease Control and Prevention

Overview

- What is an IRGT trial?
- Intraclass correlation: implications for design and analysis of IRGT trials
- Sample size and allocation to conditions
- Analytic methods for two types of IRGT trials
- Reporting requirements
- Reviews of published IRGT trials
- Future work needed

What is an IRGT trial?

- Individuals are randomly assigned to conditions (e.g. treatment, control), then interventions delivered to groups
- Potential for correlation among individuals within groups to develop over the course of the intervention
  - Therapist/facilitator effects
  - Group interaction
  - Day/time group selection

Implications of intraclass correlation

- Variance of test statistics can be larger in an IRGT trial than in a trial in which no correlation between participants is expected, due to the additional between-group variation
- Intraclass correlation coefficient (ICC)
  - Percentage of the total variation due to group membership
  - Varies between -1 and 1
- Variance inflation factor: [1+(m-1)ICC], where m is the average number of members per group
- Effective sample size=sample size/VIF

IRGT Trial Study Designs

- IRGT trials may or may not have a "baseline" measurement
- An intervention delivered in groups may be compared to:
  - Another group intervention
  - A wait-list or other individual intervention
  - Some combination of these
- The number and timing of post-intervention measurements vary

History of IRGT trials

- Recognition of issues of clustering in individually randomized trials began in psychology literature (c.f. Martindale, 1978)
- In public health, little mention of the potential for and implications of clustering in individually randomized trials until early 2000s
- At the same time, methods literature on group-randomized trials (GRTs) was growing rapidly

Hoover (2002)

- Described the potential for between-group differences in an individually randomized trial ("heterogeneous teaching subgroup effects")
- Derived the true type I error rate of tests ignoring correlation with varying numbers of groups and members per group, and this ranged from .05 to almost .50!
- Presented Satterthwaite unequal variance t-tests for studies with two group treatment conditions
- Also presented sample size formulas and an example for this approach
- Recommended mixed models to adjust for covariates

A mixed models approach

- Roberts and Roberts (2005)
  - Introduced the use of mixed models to account for clustering in IRGT trials
  - Presented formulae for differential allocation of participants to study arms based on different variances across arms
  - Examined the performance of a mixed model that allowed variance to differ across arms
  - Argued against statistical testing of the ICC to determine whether to account for it in analysis

'Partially clustered' data

- Baldwin (2011) and Bauer, Sterba & Hallfors (2008) focused on studies with clustering in some conditions, but not others
  - Recommended models that allow for between-group variance only in conditions with group treatment
  - Discouraged modeling group as a fixed effect
  - Discussed additional situations such as adjusting for baseline, outcomes with dichotomous or other distributions, and multiple treatment conditions

An interesting twist

- Andridge et al. (2014) expanded previous work to apply to IRGT trials in which participants belong to more than one group
  - SAS PROC GLIMMIX used to fit models including random effects for both groups a participant belonged to
  - Model with Kenward-Roger degrees of freedom yielded the nominal type I error rate and good power
  - A GEE approach did not perform well, but no small-sample correction was applied

Sample size for IRGT trials: two group treatments

- $$\Delta = \sqrt{\sigma_y^2 \left(\frac{(1+(m_1-1)\rho_1)}{N_1} + \frac{(1+(m_2-1)\rho_2)}{N_2}\right)(t_{\propto/2} + t_\beta)^2}$$

Δ= detectable difference

σ_y^2= between-person variance

m_1 and m_2 = members per group for conditions 1 and 2

$\rho 1$ and $\rho 2$ = ICCs for conditions 1 and 2

N_1 and N_2 = sample sizes for conditions 1 and 2

Sample size for IRGT trials: one group treatment

$$\Delta = \sqrt{\sigma_y^2 \left(\frac{(1 + (m_1 - 1)\rho_1)}{N_1} + \frac{1}{N_2}\right)(t_{\propto/2} + t_\beta)^2}$$

Δ= detectable difference

σ_y^2= between-person variance

m_1 = number of members per group for condition 1

$\rho 1$ = ICC for condition 1

N_1 = sample size for condition 1

Allocation ratio (Roberts, 2005)

$$R = \sqrt{\frac{(1+(m_1-1)\rho_1)}{(1+(m_2-1)\rho_2)}}$$

- m_1 and m_2 = average group sizes in conditions 1 and 2
- $\rho 1$ and $\rho 2$ = ICCs in conditions 1 and 2
- Ratio of variance inflation factors
- Formula reduces to the numerator in studies with only one group treatment condition

How do I choose an ICC?

- Want to pick an ICC derived from a study as similar as possible to the planned study
  - Study design
  - Duration of group interaction
  - Outcome variable
- What if such an estimate is not available?
  - Unpublished data?
  - Estimate from a cluster-randomized trial?

A few published ICCs

- Creamer, Morris, Biddle, and Elliot (1999) report ICCs from a 12-week Posttraumatic Stress Disorder (PTSD) intervention for veterans that was conducted in groups of 6-8 participants. Intraclass correlations from this study for a variety of psychosocial measures ranged from 0.04 to 0.13.
- Herzog et al., 2002 reported ICCs from a group-based smoking cessation intervention. Participants were assigned to groups based on the timing of their request for services, and groups were randomly assigned to treatments. ICC estimates of 0.32 and 0.44 were reported for group meeting attendance and smoking behavior, respectively.
- Baldwin (2011) examined ICCs for psychotherapy studies, compiling a database available on request
- Roberts & Roberts (2005) reported ICCs for a study of psychotherapy for schizophrenia, ranging from .20-.46 for schizophrenia symptom scores
- Bauer, Sterba & Hallfors (2008)- youth having academic and behavioral issues assigned to a group intervention; icc for "deviant peer bonding" was 0.06

Analysis of IRGT trials: two group treatments

- Mixed models
  - can incorporate covariates (including a baseline measurement)
  - estimates covariance parameters and can allow these to vary across conditions
- Generalized estimating equations (GEE)
  - can take correlation into account in variance estimation, but doesn't explicitly model the group variance
  - also may need a small-sample correction
- Permutation tests (also called exact tests)
  - Under the null hypothesis of no effect of the treatment, groups are 'exchangeable', meaning they could be in either treatment condition
  - P-value of observed effect can be located in a distribution of all possible effects obtained by exchanging (permuting) groups within conditions
  - Inadvisable if group sizes differ across arms
- Bayesian methods- may be more flexible in allowing incorporation of prior information on ICC and other parameters of interest

Analysis of IRGT trials: comparing a group treatment to individual treatments

- t-tests constructed using variance for each condition estimated separately
- Mixed models that allow the correlation structure to vary across arms, modeling correlation in arms with group treatment only
- Bayesian methods allowing incorporation of prior information about covariance parameters and allowing these to vary across arms

A note on degrees of freedom

- Degrees of freedom should be based on the number of independent units in the analysis
  - Groups in a GRT
  - Individuals in an RCT
  - IRGT trials-?
    - Hoover (2002) used Satterthwaite method
    - Baldwin (2011) examined Satterthwaite method and Kenward-Roger method and found little difference
    - Andridge (2014) recommended Kenward-Roger

Recommendations: Reporting Requirements

- Boutron et al. (2008) extension of CONSORT statement:
  - For each study condition, whether treatment was administered in groups or not
  - Number of groups and members per group
  - Sample size calculation, including ICC or VIF
  - ICC for all trial outcome variables (by study condition, if applicable)
  - Analytic methods used including degrees of freedom

2008 Review of published IRGT trials

Study characteristics	Number of articles	%
Journal		
American Journal of Public Health	4	11.8
Preventive Medicine	6	17.6
Health Psychology	8	23.5
Obesity	7	20.6
Addictive Behaviors	7	20.6
AIDS and Behavior	2	5.9
Number of study conditions		
Two	23	67.6
Three	8	23.5
Four	3	8.8
Number of group treatment conditions		
One	11	32.3
Two	17	50.0
Three	4	11.8
Four	2	5.9
Baseline sample size		
<100	15	44.1
100-<200	9	26.5
200-<300	4	11.8
>300	6	17.6

2008 review methods

- Results:
  - 32 of the 34 articles (94%) reported analysis at an individual level, ignoring the group entirely
  - 2 reported mixed-model analyses, and 1 reported structural equation modeling
  - Only 1 article reported appropriate analyses, 32 reported inappropriate analyses and 1 did not have enough information to judge the analytic methods

2011 Review

- Reviewed HIV/AIDS-focused GRTs and IRGT trials published in 7 journals
- Identified 25 IRGT trials
  - None reported sample size calculations taking intraclass correlation into account
  - Two reported appropriate analytic methods
  - Twenty-one reported at least one significant trial outcome

Baldwin, Murray & Shadish (2005)

- Examined 33 group psychotherapy treatments designated as empirically supported treatments by the American Psychological Association
  - None appropriately analyzed their trial data
  - Between 6 and 19 of the treatments would no longer be significant if appropriate methods used
  - Recommended compiling a database of ICCs so that corrections to prior studies could be made with better precision

Conclusions

- Although work on GRTs has been plentiful and recognition of the impact of correlation increasing, the same cannot be said for IRGT trials
- Methods work has demonstrated that ignoring the ICC, and sometimes different group sizes and different variance across conditions can inflate the type I error rate
- Sample size and analytic methods are now available for a variety of study designs, and article reviewers and journal editors should require appropriate methods

Future work needed for IRGT trials

- ICC estimate database for public health outcomes- need data from IRGT trials with a variety of study designs, intervention durations, outcome variables, etc.
- Determine impact of covariates on IRGT trial ICCs
- Sample size estimation software
- Reviews to determine whether awareness and use of proper design and analytic methods improves
- Re-analysis of IRGT trial data originally analyzed improperly (not always possible)

Questions or offers of IRGT trial data/ICCs:

Sherri L. Pals, Ph.D.
Email: sfv3@cdc.gov

References

Books:

Donner, A., & Klar, N. (2000). *Design and Analysis of Cluster Randomization Trials in Health Research*. New York: Oxford University Press.

Murray, D. M. (1998). *Design and Analysis of Group-Randomized Trials*. Oxford: Oxford University Press.

Journal Articles:

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.

Creamer M, Morris P, Biddle D, Elliot P. Treatment outcome in Australian veterans with combat-related posttraumatic stress disorder: A cause for cautious optimism? Journal of Traumatic Stress. 1999;12:545-558.

Baldwin, S. A., Bauer, D. J., Stice, E., & Rohde, P. (2011). Evaluating models for partially clustered designs. Psychological Methods, 16, 149-165. doi:10.1037/a0023464

Baldwin, S. A., Murray, D. M., & Shadish, W. R. (2005). Empirically supported treatments or Type I errors? Problems with the analysis of data from group-administered treatments. Journal of Consulting and Clinical Psychology, 73, 924-935. doi:10.1037/0022-006X.73.5.924

Baldwin, S. A., Murray, D. M., Shadish, W. R., Pals, S. L., Holland, J. M., Abramowtiz, J. S., . . . Watson, J. (2011). Intraclass correlation associated with therapists: Estimates and applications in planning psychotherapy research. Cognitive Behaviour Therapy, 40, 15-33.

Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. Multivariate Behavioral Research, 43, 210-236.

Hoover D.R. Clinical trials of behavioural interventions with heterogeneous teaching subgroup effects. Statistics in Medicine. 2002;21:1351–1364

Lee KJ, Thompson SG. The use of random effects models to allow for clustering in individually randomized trials. Clinical Trials. 2005;2:163–173.

References

Martindale C. The therapist-as-fixed-effect fallacy in psychotherapy research. Journal of Consulting and Clinical Psychology. 1978;46:1526–1530.

Pals SL, Murray DM, Alfano CM, Shadish WR, Hannan PJ, Baker WL. Individually randomized group treatment studies: Are the most frequently used analytic models correct? American Journal of Public Health. 2008; 98: 1418-1424.

Pals SL, Wiegand RE, Murray DM. Ignoring the group in group-level HIV/AIDS intervention trials: a review of reported design and analytic methods. AIDS. 2011;25(7):989-96.

Roberts, C., & Roberts, S. A. (2005). Design and analysis of clinical trials with clustering effects due to treatment. Clinical Trials, 2, 152-162.

Schnurr PP, Friedman MJ, Foy DW, et al. Randomized trial of trauma-focused group therapy for posttraumatic stress disorder: results from a department of veterans affairs cooperative study. Archives of General Psychiatry. 2003;60:481-489.

Additional Resources

- Chris Roberts' page, containing instructions for installing and using STATA Ado file clsampsi:
  - http://research.bmh.manchester.ac.uk/biostatistics/research/software/clsampsi/
- NIH page on IRGT trials:
  - https://researchmethodsresources.nih.gov/irgt.aspx