Design and Analysis of Group-Randomized Trials in Cancer: A Review of Current Practices

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

Society for Prevention Research
June 2, 2017

National Institutes of Health
Office of Disease Prevention

Co-Authors

- Sherri L. Pals, PhD, CDC
- Stephanie George, PhD, MPH, MA, ODP/NIH
- Andrey Kuzmichev, PhD, OSG/HHS
- Gabriel Lai, PhD, NCI/NIH
- Jocelyn Lee, PhD, ODP/NIH
- Ranell L. Myles, PhD, MPH, CHES, ODP/NIH
- Shakira Nelson, PhD, NCI/NIH

Group Randomized 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.
    - Also known as cluster-randomized trials.

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.
- Those connections will create a positive intraclass correlation (ICC) that reflects extra variation attributable to the group.

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

- Positive ICC will invalidate standard analytic methods.

Impact on the Analysis

- 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.
- 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
- Extra variation and limited df always reduce power.

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

Previous Reviews of the GRT Literature

- The first review was published by Donner et al. in 1990.
  - Only 19% took the ICC into account in the sample size calculations.
  - Only 50% took the ICC into account in the analysis.
- A review by Simpson et al. in 1995 reported little progress.
  - Only 19% took the ICC into account in the sample size calculations.
  - Only 57% took the ICC into account in the analysis.

- Donner A, Brown KS, Brasher P. A methodologic review of non-therapeutic intervention trials employing cluster randomization, 1979-1989. International Journal of Epidemiology. 1990;19(4):795-800.
- Simpson JM, Klar N, Donner A. Accounting for cluster randomization: a review of Primary Prevention Trials, 1990 through 1993. American Journal of Public Health. 1995;85(10):1378-83.

Previous Reviews of the GRT Literature

- A review by Varnell et al. in 2004 reported no progress, though the standards were higher than in previous reviews.
  - Only 15% took the ICC into account in the sample size calculations.
  - Only 54% always took the ICC into account in the analysis.
- A review by Murray et al. in 2008 reported mixed results, though the standards were again higher than in previous reviews.
  - Only 24% took the ICC into account in the sample size calculations.
  - Only 45% of the articles reported only analyses judged to be appropriate.

- Varnell SP, Murray DM, Janega JB, Blitstein JL. Design and analysis of group-randomized trials: a review of recent practices. American Journal of Public Health. 2004;94(3):393-9. PMC1448264.
- Murray DM, Pals SP, Blitstein JL, Alfano CM, Lehman J. Design and analysis of group-randomized trials in cancer: a review of current practices. Journal of the National Cancer Institute. 2008;100(7):483-91.

Procedures

- Systematic review of cancer-related studies published 2011-15.
  - Medline and PubMed search.
  - Studies had as their primary outcome cancer risk factors, cancer morbidity, or cancer mortality.
  - All studies were GRTs.
  - Where the paper referred to an earlier "design paper", we also reviewed that paper.
  - Each reviewer independently assessed the article on items related to design, sample size estimation, and analysis.
  - The reviewers discussed each paper as a group and any disagreements were resolved in discussion.

Findings

- 123 articles from 76 journals met the inclusion criteria.
  - 7.3% in *Preventive Medicine*
  - 4.1% in *American Journal of Preventive Medicine*
  - No more than 3.3% in any other single journal
- 24.6 GRT papers per year (2011-15)
  - vs. 15.0 per year (2002-06) in Murray et al.
  - vs. 11.6 per year (1998-02) in Varnell et al.
  - vs. 5.3 per year (1990-93) in Simpson et al.
- 39 background "design" papers.

Findings

- 24% of the articles made no mention of sample size calculations.
  - Down from 47% in 2008
- 53% reported appropriate methods for sample size.
  - Up from 24% in 2008

Table 1. Analytic methods frequently used in group-randomized trials (GRTs) and the conditions under which their use is appropriate.

Method	Appropriate Application
Mixed-model methods	
ANOVA/ANCOVA[a]	One time point in the analysis
Repeated measures ANOVA/ANCOVA	Two time points in the analysis
Random coefficients approach	Three or more time points in the analysis
Generalized Estimating Equations	
With correction for limited df[b]	Fewer than 40 groups included in analysis
With no correction for limited df	40 or more groups included in analysis
Cox regression	
With shared frailty	Any GRT with a time-to-event outcome
Without shared frailty	Not appropriate for group-randomized trials
Two-stage Methods (analysis on group means or other summary statistic)	Applied at the level of the unit of assignment

[a] ANOVA: analysis of variance; ANCOVA: analysis of covariance

[b] df: degrees of freedom

Table 1. Analytic methods frequently used in group-randomized trials (GRTs) and the conditions under which their use is appropriate.

Method	Appropriate Application
Post-hoc correction based on external estimates of intraclass correlation	Validity depends on validity of external estimates of intraclass correlation
Analysis at subgroup level, ignoring group-level intraclass correlation	Not appropriate for GRTs
Analysis at individual level, ignoring group-level intraclass correlation	Not appropriate for GRTs
Analysis at individual level, modeling group as a fixed effect	Not appropriate for GRTs

[a] ANOVA: analysis of variance; ANCOVA: analysis of covariance
[b] df: degrees of freedom

Table 3. Characteristics of 123 articles which reported the results of group-randomized trials in cancer research in selected peer-reviewed journals during the period 2011-2015, inclusive.

Characteristic	N	%
Number of Study Conditions		
Two	109	88.6
Three	9	7.3
Four or more	5	4.1
Design		
Cohort	94	76.4
Cross-sectional	26	21.1
Combination of Cohort and Cross-sectional	3	2.4
Matching or Stratification in Design		
Matching only	16	13.0
Stratification only	46	37.4
Constrained Randomization only	2	1.6
Matching and Stratification	3	2.4
Randomization without Matching or Stratification	56	45.5

Table 3. Characteristics of 123 articles which reported the results of group-randomized trials in cancer research in selected peer-reviewed journals during the period 2011-2015, inclusive.

Characteristic	N	%
Type of Group		
Churches	6	4.9
Communities, Neighborhoods or Community Groups	15	12.2
Families	4	3.3
Housing Projects or Apartment Buildings	1	0.8
Clinicians, Provider Groups, Hospitals	65	52.8
Schools, Classes, Day Care Centers	24	19.5
Time period[1]	4	3.3
Worksites	4	3.3
Number of Groups per Condition in the Analysis		
1 Group	0	0.0
2-5 Groups	3	2.4
6-8 Groups	9	7.3
9-12 Groups	16	13.0
13-24 Groups	31	25.2
> 25 Groups	58	47.2
Variable	1	0.8
not reported	5	4.1

Table 3. Characteristics of 123 articles which reported the results of group-randomized trials in cancer research in selected peer-reviewed journals during the period 2011-2015, inclusive.

Characteristic	N	%
Number of Members per Group in the Analysis		
<10 Members	30	24.4
10-49 Members	44	35.8
50-99 Members	19	15.4
>100 Members	25	20.3
not reported	5	4.1
Number of Time Points in the Analysis		
1 Time point	94	76.4
2 Time points	21	17.1
3-9 Time points	8	6.5
Focus of Study		
Primary Prevention	45	36.6
Secondary Prevention	54	43.9
Tertiary Prevention	24	19.5

Table 3. Characteristics of 123 articles which reported the results of group-randomized trials in cancer research in selected peer-reviewed journals during the period 2011-2015, inclusive.

Characteristic	N	%
Primary Outcome Variables		
Alcohol Use	3	2.4
Delivery of Health Services	22	17.9
Dietary Variables	9	7.3
Fatigue	0	0.0
Incidence of Cancer	4	3.3
Knowledge of Cancer or Attitudes Regarding Cancer	10	8.1
Lymphedema	0	0.0
Mortality from Cancer	1	0.8
Neuropathy	0	0.0
Pain	3	2.4
Physical Activity	5	4.1
Quality of Life	6	4.9
Screening	33	26.8
Sun Protection	3	2.4
Tobacco Use	10	8.1
Weight	1	0.8
Other	13	10.6

Table 4. Distribution of analytic methods in 123 articles reporting on group-randomized trials in cancer research published in peer-reviewed journals during the period 2011-2015, inclusive.

Criteria	N	%	N	%
Articles reporting only appropriate methods	63	51.2		
Mixed-model ANOVA or ANCOVA with 1 time point			39	56.5
Mixed-model repeated measures with 2 time points			7	10.1
Random coefficients model with >2 time points			2	2.9
Generalized Estimating Equations with ≥40 groups			9	13.0
Cox regression with adjustment for the unit of assignment			4	5.8
Two-stage analysis			6	8.7
Other			2	2.9

Table 4. Distribution of analytic methods in 123 articles reporting on group-randomized trials in cancer research published in peer-reviewed journals during the period 2011-2015, inclusive.

Criteria	N	%	N	%
Articles reporting both appropriate and inappropriate methods	17	13.8		
Appropriate Methods				
Mixed-model ANOVA or ANCOVA with one time point			11	64.7
Mixed-model repeated measures with two time points			1	5.9
Random coefficients model with more than two time points			0	0.0
Generalized Estimating Equations with 40 or more groups			2	11.8
Cox regression with shared frailty for the unit of assignment			3	17.6
Two-stage analysis			0	0.0
Other			0	0.0
Inappropriate Methods				
Analysis at an individual level, ignoring group-level ICC			16	94.1
Analysis at a subgroup level, ignoring group-level ICC			0	0.0
Analysis with group as a fixed effect			0	0.0
Mixed-model repeated measures, > 2 time points			0	0.0
Asymptotically robust method with fewer than 40 groups			1	5.9
Individual-level analysis with post-hoc correction			0	0.0
Other			0	0.0

Table 4. Distribution of analytic methods in 123 articles reporting on group-randomized trials in cancer research published in peer-reviewed journals during the period 2011-2015, inclusive.

Criteria	N	%	N	%
Articles reporting only inappropriate methods	37	30.1		
Analysis at an individual level, ignoring group-level ICC			18	45.0
Analysis at a subgroup level, ignoring group-level ICC			7	17.5
Analysis with group as a fixed effect			2	5.0
Mixed-model repeated measures, > 2 time points			3	7.5
Asymptotically robust method with fewer than 40 groups			8	20.0
Individual-level analysis with post-hoc correction			1	2.5
Other			1	2.5
Not enough information provided	6	4.9		

Summary

- We report improvements in the methods used to design and analyze GRTs in three areas:
  - The proportion of studies that documented that their sample size calculations had been done correctly increased from 24% to 53%.
  - The proportion of studies that used only analyses deemed appropriate increased from 45% to 51%.
  - The proportion of studies that failed to report enough information to classify their analytic methods declined from 12% to 5%.
- Even so, there is ample room for further improvement.
  - Many studies continue to analyze data at the individual level, ignoring the group-randomization.
  - Others analyze at a subgroup level, ignoring the group-level ICC.
  - Others use asymptotic methods when asymptotic conditions do not apply.

Efforts to Address This Problem

The NIH Office of Disease Prevention

- Free 7-part online course on GRTs available on the ODP website.
- Web-based search tool for Scientific Review Officers based on the Prevention Research Expertise Survey (PRES) database
- Standard language in future FOAs that may support clinical trials will alert investigators to these issues.
- Modifications to the review criteria in these FOAs will alert investigators that these issues will be considered in review.
- Drafting GRT training materials that may be provided to reviewers and SROs for future applications that involve clinical trials.

Pragmatic and Group-Randomized Trials in Public Health and Medicine

Visit **https://prevention.nih.gov/grt** to:

- Provide feedback on this series
- Download the slides, references, and suggested activities
- View any of the modules in this series

Send questions to:
GRT@mail.nih.gov

National Institutes of Health
Office of Disease Prevention