Assessing the most powerful analysis method for school-based intervention studies

Jessica B. Janega  
*University of Memphis, jjanega@mail.psyc.memphis.edu*

David M. Murray  
*University of Memphis, d.murray@mail.psyc.memphis.edu*

Sherri P. Varnell  
*Division of HIV/AIDS Prevention, sfv3@cdc.gov*

Jonathan L. Blitstein  
*University of Memphis, jblitstn@memphis.edu*

Amanda Birnbaum  
*Montclair State University, birnbauma@montclair.edu*

See next page for additional authors.

MSU Digital Commons Citation  
Janega, Jessica B.; Murray, David M.; Varnell, Sherri P.; Blitstein, Jonathan L.; Birnbaum, Amanda; and Lytle, Leslie A., "Assessing the most powerful analysis method for school-based intervention studies" (2004). *Department of Public Health Scholarship and Creative Works*. 70.  
https://digitalcommons.montclair.edu/public-health-facpubs/70

This Article is brought to you for free and open access by the Department of Public Health at Montclair State University Digital Commons. It has been accepted for inclusion in Department of Public Health Scholarship and Creative Works by an authorized administrator of Montclair State University Digital Commons. For more information, please contact digitalcommons@montclair.edu.
Assessing the most powerful analysis method for school-based intervention studies with alcohol, tobacco, and other drug outcomes

Jessica B. Janega\textsuperscript{a,1}, David M. Murray\textsuperscript{a,*}, Sherri P. Varnell\textsuperscript{b,2}, Jonathan L. Blitstein\textsuperscript{a,1}, Amanda S. Birnbaum\textsuperscript{c,3}, Leslie A. Lytle\textsuperscript{c,4}

\textsuperscript{a}Department of Psychology, The University of Memphis, 202 Psychology Building, 3693 Norriswood, Memphis, TN 38152-3230, USA
\textsuperscript{b}Contractor Support to Division of HIV/AIDS Prevention, Epidemiology Branch, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop E-45, Atlanta, GA, USA
\textsuperscript{c}Division of Epidemiology, School of Public Health, University of Minnesota, 1300 South Second Street, Suite 300, Minneapolis, MN, USA

Abstract

This article compares four mixed-model analyses valid for group-randomized trials (GRTs) involving a nested cohort design with a single pretest and a single posttest, the most common design used in GRTs. This study makes estimates of intraclass correlations (ICCs) available to investigators planning GRTs with alcohol, tobacco, and other drug measures as the outcomes of interest. It also provides formulae demonstrating the potential benefits to the standard error of the intervention effect of both adjustments for fixed and time-varying covariates, as well as correlations over time. These estimates will allow other researchers using these variables to plan their studies by performing a priori power analyses for any of four common analytic options.

\textcopyright{} 2004 Elsevier Ltd. All rights reserved.

Keywords: Intraclass correlation; Group-randomized trial; Power; School based; ATOD prevention
1. Introduction

Group-randomized trials (GRTs) are comparative studies in which the units of assignment are identifiable groups and the units of observation are members of those groups (Murray, 1998). GRTs are one of the best comparative designs available when investigators wish to explore the effects of interventions delivered at the group level. They are widely used in public health, education, and sociology (Donner & Klar, 2000; Murray, 1998). Such trials employ different units of assignment and observation; this poses a number of design and analytic problems absent when individuals are randomized to conditions (Murray, 1998). One of these problems is that observations taken from members of the same identifiable group are likely to have something in common, due to commonality in selection, exposure, or mutual interaction (Kish, 1965). This commonality is indexed by an intraclass correlation (ICC) and reflects a component of variance attributable to the groups, $r_g^2$, in addition to the usual variation attributable to the members within those groups, $r_m^2$.

For a simple mixed-model analysis of variance (ANOVA), the total variance for the dependent variable $y$ is $\sigma_y^2 = \sigma_m^2 + \sigma_g^2$. The ICC from that analysis is calculated as:

$$\text{ICC}_{m:g:c} = \frac{r_g^2}{\sigma_m^2 + r_g^2} \tag{1}$$

Here, m:g:c reflects nesting of members within groups within conditions (Murray, 1998).

When the ICC is ignored in the analysis of the intervention effect in a GRT, simulation studies have shown that the Type I error rate is inflated (Murray, Hannan, & Baker, 1996; Murray & Wolfinger, 1994; Zucker, 1990). However, when the ICC is properly reflected in the analysis, power may be limited, due in part to the extra variation attributable to the groups and to the limited degrees of freedom (df) available to estimate that extra variation (Cornfield, 1978). Given that several valid analyses are available for GRTs, it is important to know how they compare in terms of power and whether it is possible to determine, in advance, when one method is more powerful than another.

We have long assumed that ICC estimates from baseline data were appropriate to estimate power even when the analysis would ultimately involve follow-up data. However, recent work suggests that ICCs estimated from baseline data may not be good proxies for ICCs observed at follow-up (Janega et al., in press). Unfortunately, there are very few papers that have published ICCs estimated from follow-up data (Feng et al., 1999; Murray et al., 2002; Murray, Clark, & Wagenaar, 2000). To help address this problem, this paper provides ICC estimates using follow-up data for alcohol, tobacco, and other drug endpoints from the Teens Eating for Energy and Nutrition at School (TEENS) study (Lytle & Perry, 2001).

The TEENS study employed a nested cohort design with a single pretest and a single posttest. This is the most common design used in GRTs and several valid analyses exist. The simplest is the mixed-model ANOVA which uses posttest data only. In this analysis, the intervention effect, $\Delta$, is the simple unadjusted difference between the intervention and
control condition means. The mixed-model analysis of covariance (ANCOVA) is another option. Here, we analyze the posttest data with regression adjustment for baseline covariates, including the baseline value of the dependent variable. The $\Delta$ is the simple difference between the condition means at posttest after adjustment for any difference attributable to baseline covariates. The mixed-model repeated-measures ANOVA is a third option. The mixed-model repeated-measures ANOVA models time explicitly, and $\Delta$ is the net difference among the intervention and control condition pretest and posttest means. The fourth option is a mixed-model repeated-measures ANCOVA, which models time explicitly and makes adjustments for time-varying covariates. The $\Delta$ is the net difference among the adjusted intervention and control condition pretest and posttest means.

Given a pretest–posttest nested cohort design with two study conditions, all four of these models have 1 $df$ for the numerator of the test of $\Delta$; as a result, the most powerful analysis for a given $\Delta$ will be the analysis with the smallest standard error, $\sigma_\Delta$. The $\sigma_\Delta$ is affected by several factors, including the ICC, variance reduction due to regression adjustment for covariates, and variance reduction due to over-time correlation.

For the ANOVA and ANCOVA, a general formula for $\sigma_\Delta$ is provided by Murray (1998) as:

$$
\sigma_\Delta = \sqrt{2 \left( \frac{\sigma_y^2(1 - \text{ICC}_{m:g:c}) \theta_m + m(\sigma_y^2(\text{ICC}_{m:g:c})) \theta_g}{mg} \right)}
$$

(2)

The total random variance is $\sigma_y^2 = \sigma_m^2 + \sigma_g^2$, the member component of variance is $\sigma_m^2 = \sigma_y^2 (1 - \text{ICC}_{m:g:c})$, and the group component of variance is $\sigma_g^2 = \sigma_y^2 (\text{ICC}_{m:g:c})$. $\theta_m$ and $\theta_g$ reflect the change in those components due to regression adjustment for covariates, if any; they are defined as the ratio of the adjusted and unadjusted components of variance. They equal 1 if there is no regression adjustment.

For the repeated-measures ANOVA and ANCOVA, a general formula for $\sigma_\Delta$ is provided by Murray (1998) as:

$$
\sigma_\Delta = \sqrt{2 \times 2 \left( \frac{\sigma_y^2(1 - \text{ICC}_{m:g:c})(1 - r_{yy}(m)) \theta_m + m(\sigma_y^2(\text{ICC}_{m:g:c}))(1 - r_{yy}(g)) \theta_g}{mg} \right)}
$$

(3)

Here, most of the terms are defined as in Eq. (2), but they will have different values because they are estimated from the pretest and posttest data, rather than just from the posttest data. New in Eq. (2) are the $r_{yy}(g)$ and $r_{yy}(m)$, which are correlations over time for groups and members.

We will report estimates of ICC$_{m:g:c}$, $\sigma_g^2$, $\sigma_m^2$, $r_{yy}(g)$, $r_{yy}(m)$, $\theta_g$, $\theta_m$, and $\sigma_\Delta$ as appropriate for each of these four analyses, drawing either from the nonrepeated-measures analysis of the TEENS posttest data or from the repeated-measures analysis of the TEENS pretest and posttest data. We will illustrate how these factors contribute to power and provide examples of how to use these estimates to plan future GRTs.
2. Materials and methods

2.1. The TEENS study design and survey procedures

The TEENS study recruited 16 middle schools from Minneapolis-St. Paul (Lytle & Perry, 2001). TEENS implemented school-, classroom-, and family-level interventions to reduce cancer-related dietary risk behaviors and increase health-promoting behaviors in seventh and eighth graders. TEENS focused on a lower-income population and included only districts in which at least 20% of the students were eligible for free or reduced-price lunches. Participating schools were also required to have both seventh and eighth graders in the same building and to enroll at least 30 students per grade. In 14 districts, 33 schools were eligible and of those, 20 schools in 9 districts agreed to participate. Schools that refused to participate did so due to time constraints, personnel changes, and lack of interest in the school food-environment component of the intervention. Of the 20 schools that agreed to participate, 1 was used in a pilot study and 3 others were excluded due to scheduling conflicts. The remaining 16 schools were randomly assigned from matched pairs to intervention and control conditions. Evaluation of TEENS included student surveys at baseline and follow-up and 24-h dietary recall interviews from a random sample of those students.

Baseline data were gathered in the fall semester of 1998. Trained staff members administered baseline surveys in a required seventh-grade class. Staff members noted absences and made one follow-up visit to reach absent students. Of the 4050 eligible seventh graders, 95 (2.3%) were absent from school during two survey attempts, 77 (1.9%) were excluded due to parental or student refusal, and 3878 (95.9%) completed the in-class survey.

The research team collected follow-up survey data in the spring semester of 2000 using the same procedures. Of the 3878 seventh graders who provided baseline survey data, 3010 (77.6%) provided follow-up survey data as eighth graders. Thus, the loss to follow-up rate for the survey was 22.4%. An analysis of the baseline data showed that students who were lost to follow-up were more likely to report minority background, living with one parent, and participation in the free or reduced-price lunch program; they were less likely to report two parents working full time, and parents with higher levels of educational attainment.

2.2. Variables of interest and their measures

The Minnesota Smoking Index and items from the Monitoring the Future (MTF) study determined alcohol, tobacco, and drug (ATOD) use (Johnston, O’Malley, & Bachman, 1996; Pechacek, Fox, Murray, & Luepker, 1984). The Minnesota Smoking Index has shown good agreement with biochemical markers for tobacco exposure and provides measures of daily and weekly smoking. One MTF item, “How frequently have you smoked cigarettes during the past 30 days?” assessed monthly cigarette smoking prevalence (1 = any, 0 = none); test–retest reliability was moderate ($\kappa=.67$). Another MTF item, “During the last 30 days, how
may times have you had alcohol to drink (including beer, wine, and liquor)?” assessed any prevalent alcohol use in the past 30 days (1 = any, 0 = none); test–retest reliability was moderate ($\kappa = .50$). Another MTF item assessed binge drinking, “Think back over the last 2 weeks. How many times have you had five or more drinks in a row?” (1 = any, 0 = none); test–retest reliability was moderate ($\kappa = .52$). Other MTF items assessed marijuana use, “During the last 30 days, how may times have you: Used marijuana?” (1 = any, 0 = none), and inhalant use, “During the last 30 days, how may times have you: Sniffed glue, gas, sprays, or anything else like that to get high?” (1 = any, 0 = none); test–retest reliability was moderate for marijuana ($\kappa = .60$) and inhalants ($\kappa = .47$). A composite was used for any ATOD use in the last 30 days (1 = any, 0 = none).

Violent behavior was assessed using five items with a common stem, “In the last 12 months, how often did you...” The items were: “Carry a weapon such as a gun, knife or club” (Kahn, Kinchen, Williams, Ross, Lowry, Hill, Grunbaum, Blumson, Collins and Kolbe, 1998); “Hit or beat up someone” (Minnesota Department of Education, 1989); “Take part in a fight where a group of your friends fought another group”, “Hurt someone badly enough to need bandages or a doctor”, and “Use a knife, gun, or other weapon to get something from a person” (Carolina Population Center, 1999). Response categories were: never, 1–3 times, 4–7 times, 8–11 times, and 12 or more times. We assigned the midpoint for each response category (for the last category, we assigned 14) and summed across items. The scale ranged from 0 to 70, with higher scores indicating more violent behavior. Test–retest reliability for this measure was good (Spearman $r = .76$); internal consistency at baseline was also good (Cronbach’s $\alpha = .73$).

The Center for Epidemiologic Studies—Depression Scale (CES-D; Radloff, 1977) assessed the frequency of occurrence of depressive symptoms. This 20-item scale has been used with both adult and adolescent populations (Doerfler, Felner, Rowlinson, Raley, & Evans, 1988; Radloff, 1977). Test–retest correlation for the CES-D scale items is good (Spearman $r = .82$), as is the internal consistency of the scale in the baseline administration (Cronbach’s $\alpha = .86$).

Demographic variables included gender (male or female), race (White, Black, Hispanic, Asian, Native American, Mixed, or Other), household structure (lives with two parents or another arrangement), parents’ education attainment (both parents have at least some high school education, one parent has vocational school or some college training, one parent has completed college, both parents have at least completed college, data are missing or student does not know parental education status, or other), parents’ employment status (both parents work full time, one parent works full time, or other), and eligibility for the free or reduced-price lunch program (yes or no; Birnbaum et al., 2002).

2.3. Analysis methods

To be able to compare the four analytic models, they had to be run on the same individuals. As a result, individuals who did not provide complete information at baseline and follow-up were excluded from the analysis. We then applied the four analyses to each of the dependent variables using either PROC MIXED or the GLIMMIX macro in
SAS Version 8.2 (SAS Institute, 1999). MIXED implements the General Linear Mixed Model and is appropriate for normally distributed data with multiple random effects while the GLIMMIX macro implements the Generalized Linear Mixed Model and is appropriate for data with multiple random effects and a non-Gaussian residual error distribution (Littell, Milliken, Stroup, & Wolfinger, 1996; Murray, 1998). PROC MIXED was used for the analysis of the normally distributed dependent variable depressive symptoms. The GLIMMIX macro was used with a log link and Poisson error function for the count variable violent behavior and with a logit link and binomial error function for the dichotomous variables weekly smoking, monthly smoking, alcohol use, binge drinking, marijuana use, inhalant use, and ATOD use.

The mixed-model ANOVA is performed on posttest data only. For each dependent variable, the ANOVA estimates the group and member components of variance, σ^2 and σ^2_m; the sum of these two components is the total variance for the dependent variable, σ^2_y. We estimated the ICC_m:g:c using Eq. (1) and σ_Δ using Eq. (2).

In the mixed-model ANCOVA, we adjusted for the baseline measurement of the dependent variable, gender, ethnicity, age at baseline, whether or not the child reported living with two parents home at baseline, whether or not the child was eligible for the free or reduced-price lunch program at baseline, parental employment status at baseline, and parental education attainment at baseline. For each dependent variable, the ANCOVA estimates σ^2_m and σ^2_g adjusted for the covariates. We calculated θ_m and θ_g as the ratio of the adjusted and unadjusted σ^2_m and σ^2_g drawing on the results of the mixed-model ANOVA for the unadjusted estimates (Murray, 1998). We estimated σ_Δ using Eq. (2).

The mixed-model repeated-measures ANOVA is performed on the pretest and posttest data. This analysis models time explicitly and estimates σ^2_m, σ^2_g, r_yy(g), and r_yy(m). The estimates of σ^2_m and σ^2_g usually differ from those obtained in the non-repeated-measures mixed-model ANOVA because they are based on both the pretest and posttest data, rather than on the posttest data alone. We estimated σ_Δ using Eq. (3).

In the mixed-model repeated-measures ANCOVA, we adjusted for several time-varying covariates: age, whether or not the child reported living with two parents, whether or not the child was eligible for the free or reduced-price lunch program, parental employment status, and parental educational attainment. The mixed-model repeated-measures ANCOVA is also performed on the pretest and posttest data and models time explicitly. We calculated θ_m and θ_g drawing on the mixed-model repeated-measures ANOVA for the unadjusted estimates (Murray, 1998). We estimated σ_Δ using Eq. (3).

3. Results

Of the 3878 students who provided baseline survey data, 3009 were included in the analyses reported here and 869 (22.4%) were excluded; of the exclusions, 868 did not complete the follow-up survey and 1 student was missing a value for at least one variable of interest. Compared to the students included in the analyses reported here, students excluded
from the analyses reported higher levels of ATOD use, violent behavior and depressive symptoms; were less likely to report living with two parents, having both parents employed full time, and having parents with higher levels of educational attainment; and were more likely to report eligibility for the free or reduced-price lunch program, and self-identification with a minority ethnic group.

For each analysis and each dependent variable, Table 1 presents the estimate of the $\text{ICC}_{m:g:c}$, $\sigma^2_g$, $\sigma^2_m$, $r_{yy(g)}$, $r_{yy(m)}$, $\theta_g$, $\theta_m$, and $\sigma_\Delta$. For example, the mean percentage of alcohol use during the previous 30 days in the sample of seventh graders was 29.7%; that rate is somewhat higher than the rate of 23% reported for 1998 for eighth graders in the MTF survey (http://monitoringthefuture.org/data/01data.html#2001data-drugs). The $\text{ICC}_{m:g:c}$ was estimated as .0111, which is similar to values reported previously for the ICC for monthly alcohol use in a school-based GRT (Murray et al., 2000; Murray & Hannan, 1990; Murray & Short, 1996). Regression adjustment for covariates was most helpful at the school level and then largely in the ANCOVA, as reflected in the fractional value for the estimate of $\theta_g$. Alcohol use was correlated over time, although much more so at the level of the school than the student. For alcohol use, the ANCOVA and the repeated-measures ANCOVA both produced the lowest value for $\sigma_\Delta$, indicating that either analysis would be appropriate for this variable. Of the remaining eight variables in Table 1, the mixed-model ANCOVA produced the lowest $\sigma_\Delta$ for three of them, the mixed-model ANOVA produced the lowest $\sigma_\Delta$ for two variables, and the mixed-model repeated-measures ANCOVA produced the lowest $\sigma_\Delta$ for two variables.

4. Discussion

4.1. Comparison of the four models

Previous research with dietary variables from the TEENS study suggested that the mixed-model ANCOVA might provide the lowest $\sigma_\Delta$ more consistently than any of the other three models under consideration here (Janega et al., in press). With these data, however, that was not always the case. The mixed-model ANCOVA provided the lowest $\sigma_\Delta$ for only three of the nine variables: binge drinking behavior, marijuana use, and inhalant use. The mixed-model ANCOVA and the mixed-model repeated-measures ANCOVA provided the same $\sigma_\Delta$ for one variable, alcohol use. The mixed-model repeated-measures ANCOVA was the best model for two other variables, violent behavior and depressive symptoms. The mixed-model ANOVA provided the lowest $\sigma_\Delta$ for the final three variables, weekly smoking, monthly smoking, and the combined ATOD variable. In other words, the variables were split almost evenly among the mixed-model ANCOVA, ANOVA, and repeated-measures ANCOVA. For this reason, it is not appropriate to recommend one model over another for these variables in this dataset. As such, this paper does not replicate our previous findings with the TEENS dietary variables (Janega et al., in press); instead, these findings underscore the need to publish estimates for the variables of interest for the analysis under consideration.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (%)</th>
<th>Model&lt;sup&gt;a&lt;/sup&gt;</th>
<th>ICC&lt;sub&gt;m,g,c&lt;/sub&gt;</th>
<th>( \sigma^2_g )</th>
<th>( \sigma^2_m )</th>
<th>( r_{j,j(g)} )</th>
<th>( r_{j,j(m)} )</th>
<th>( \theta_g )</th>
<th>( \theta_m )</th>
<th>( \sigma_\Delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly cigarette smoking</td>
<td>9.24</td>
<td>ANOVA</td>
<td>0.0072</td>
<td>0.0006</td>
<td>0.0819</td>
<td>1.1607</td>
<td>0.9882</td>
<td>0.0168</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANCOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RM ANOVA</td>
<td>0.1118</td>
<td>0.0040</td>
<td>0.0585</td>
<td>0.0168</td>
<td>0.0221</td>
<td>0.0168</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RM ANCOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly cigarette smoking</td>
<td>14.9</td>
<td>ANOVA</td>
<td>0.0059</td>
<td>0.0077</td>
<td>0.1256</td>
<td>1.0620</td>
<td>1.0015</td>
<td>0.0211</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANCOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RM ANOVA</td>
<td>0.0077</td>
<td>0.0060</td>
<td>0.0882</td>
<td>0.0198</td>
<td>0.0237</td>
<td>0.0188</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RM ANCOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol use</td>
<td>29.7</td>
<td>ANOVA</td>
<td>0.0111</td>
<td>0.0234</td>
<td>0.2080</td>
<td>0.0077</td>
<td>0.0006</td>
<td>0.0268</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANCOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RM ANOVA</td>
<td>0.0042</td>
<td>0.0177</td>
<td>0.1614</td>
<td>0.0954</td>
<td>0.0607</td>
<td>0.0223</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RM ANCOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binge drinking behavior</td>
<td>10.7</td>
<td>ANOVA</td>
<td>0.0023</td>
<td>0.0002</td>
<td>0.0946</td>
<td>0.1107</td>
<td>1.0102</td>
<td>0.0266</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANCOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RM ANOVA</td>
<td>0.0002</td>
<td>0.0003</td>
<td>0.0755</td>
<td>0.9437</td>
<td>0.0262</td>
<td>0.0144</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RM ANCOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marijuana use</td>
<td>11.9</td>
<td>ANOVA</td>
<td>0.0073</td>
<td>0.0008</td>
<td>0.1035</td>
<td>0.6529</td>
<td>0.6529</td>
<td>0.0370</td>
<td>0.0284</td>
<td>0.0152</td>
</tr>
</tbody>
</table>
Inhalant use 5.29

<table>
<thead>
<tr>
<th></th>
<th>ANCOVA</th>
<th>RM ANOVA</th>
<th>ANCOVA</th>
<th>RM ANOVA</th>
<th>ANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0069</td>
<td>0.0008</td>
<td>0.0774</td>
<td>0.3349</td>
<td>0.0376</td>
</tr>
<tr>
<td></td>
<td>0.5321</td>
<td>1.0187</td>
<td>0.0247</td>
<td>0.0066</td>
<td>0.0128</td>
</tr>
</tbody>
</table>

ATOD 34.5

<table>
<thead>
<tr>
<th></th>
<th>ANCOVA</th>
<th>RM ANOVA</th>
<th>ANCOVA</th>
<th>RM ANOVA</th>
<th>ANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0059</td>
<td>0.0013</td>
<td>0.2255</td>
<td>1.2030</td>
<td>0.0252</td>
</tr>
<tr>
<td></td>
<td>0.0071</td>
<td>0.0011</td>
<td>0.1702</td>
<td>0.0336</td>
<td>0.0310</td>
</tr>
</tbody>
</table>

Violent behavior 6.44

<table>
<thead>
<tr>
<th></th>
<th>ANCOVA</th>
<th>RM ANOVA</th>
<th>ANCOVA</th>
<th>RM ANOVA</th>
<th>ANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0171</td>
<td>0.1450</td>
<td>8.3313</td>
<td>0.8140</td>
<td>0.2183</td>
</tr>
<tr>
<td></td>
<td>0.00254</td>
<td>0.1159</td>
<td>20.6252</td>
<td>0.4257</td>
<td>0.1547</td>
</tr>
</tbody>
</table>

CES-D depressive symptoms 14.2

<table>
<thead>
<tr>
<th></th>
<th>ANCOVA</th>
<th>RM ANOVA</th>
<th>ANCOVA</th>
<th>RM ANOVA</th>
<th>ANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0186</td>
<td>2.2714</td>
<td>119.8700</td>
<td>0.5172</td>
<td>0.1435</td>
</tr>
<tr>
<td></td>
<td>0.0012</td>
<td>2.2297</td>
<td>103.6814</td>
<td>0.4044</td>
<td>0.4154</td>
</tr>
</tbody>
</table>

\* Mixed-models coded as follows: ANOVA = analysis of variance, ANCOVA = analysis of covariance, RM ANOVA = repeated measures analysis of variance, RM ANCOVA = repeated measures analysis of covariance.
4.2. Power analysis

As an example of how a power analysis might be conducted using these results, consider a study designed to decrease marijuana use among junior high school students. Assume a pretest–posttest control group design with schools randomized to conditions as in the design employed in TEENS. The study could be planned based on any of the four analyses considered thus far. Based on the estimates in Table 1, we would choose a mixed-model ANCOVA, as that has the lowest $\sigma^2$. We estimate the detectable difference using the following equation (Murray, 1998):

$$\Delta = \sqrt{\sigma^2 (t_{\text{critical}:z/2} + t_{\text{critical}:\beta})^2}$$

(4)

After substituting Eq. (2) for $\sigma^2$,

$$\Delta = \sqrt{2 \left( \frac{(1 - \text{ICC}_{m:g:c}) \theta_m + m(\sigma^2_{Y}(\text{ICC}_{m:g:c})) \theta_g}{mg} \right) (t_{\text{critical}:z/2} + t_{\text{critical}:\beta})^2}$$

(5)

Degrees of freedom for the analysis are equal to $c(g - 1)$ where $c$ represents the number of conditions and $g$ represents the number of groups per condition.

If we plan a study with 100 students per school and 10 schools per condition, $df=2(10 - 1)=18$. Using the nominal two-tailed Type I error rate of .05 and 80% power, the critical values for $t$ are $t_{\text{critical}:z/2}=2.101$ and $t_{\text{critical}:\beta}=0.862$. Using the estimates from Table 1 for marijuana use from the mixed-model ANCOVA for the other parameters:

$$\Delta = \sqrt{2 \left( \frac{(1 - \text{ICC}_{m:g:c}) \theta_m + m(\sigma^2_{Y}(\text{ICC}_{m:g:c})) \theta_g}{mg} \right) (2.101 + 0.862)^2}$$

$$= 0.0545$$

(6)

In this example, there is 80% power to detect a change in marijuana use of 0.0545, or 5.45%. Stated another way, if the rate of use in the control group is 10%, we would have 80% power to detect a rate as low as 4.55% in the intervention group ($10\% - 5.45\% = 4.55\%$). Similar calculations using the estimates from the mixed-model ANOVA result in a detectable difference of 5.62%, which is a 3.12% increase in the size of the detectable difference from the mixed-model ANCOVA. The mixed-model repeated-measures ANOVA results in a detectable difference of 6.23%, a 14.31% increase in detectable difference from the mixed-model ANCOVA. Finally, the mixed-model repeated-measures ANCOVA results in a detectable difference of 5.89%, an 8.07% increase in detectable difference. This example illustrates the potential savings in detectable difference that good estimates can provide in planning a study.
4.3. Comparison of those excluded and included in the analyses

Students who were excluded from the analyses at baseline were significantly different from those who were included. Although this could limit generalizability to some extent, it does not affect the validity of the comparisons among the four analytic models because those models were run on the same participants using variables measured in the same way.

4.4. Recommendations

This study makes ICCs and other important parameter estimates available to investigators planning GRTs with ATOD measures as the outcomes of interest. It also provides formulae demonstrating the potential benefits to $\sigma_\Delta$ of both adjustments for fixed and time-varying covariates, as well as correlations over time. These estimates will allow other researchers using these variables to plan their studies by performing a priori power analyses for any of the different analysis options, and weighing the potential benefits, to choose the most appropriate analysis.

When researchers have access to estimates of over-time correlations and potential adjustments for both time-varying and fixed covariates, two main guidelines can be useful in planning future studies. First, correlations over time must be about .5 or larger to be beneficial, as they must compensate for the additional 2 in the numerator of $\sigma_\Delta$. When estimates of over-time correlation do not exceed .5, there is generally no benefit to power of selecting a mixed-model repeated-measures analysis. Second, in order for adjustments for covariates to be successful in reducing variance, $\theta$, as defined below, should be less than 1:

\[ \theta_m = \frac{\text{adjusted } \sigma^2_m}{\text{unadjusted } \sigma^2_m} \quad \text{and} \quad \theta_g = \frac{\text{adjusted } \sigma^2_g}{\text{unadjusted } \sigma^2_g} \]

(7)

It is most important for investigators to have estimates like those in Table 1 when planning GRTs. This paper provides such estimates from the alcohol, tobacco, and other drug variables from the TEENS data set, and demonstrates how to use these estimates in power calculations. Future studies should continue to publish estimates so that other investigators may choose the most appropriate analysis for their study.

Acknowledgements

This research was supported by National Cancer Institutes grant no. CA71943-011A1.

References


