Communication Studies Faculty Articles and Research

Communication Studies

2013

Random Assignment of Schools to Groups in the Drug Resistance Strategies Rural Project: Some New Methodological Twists

John W. Graham
Pennsylvania State University

Jonathan Pettigrew
Pennsylvania State University

Michelle Miller-Day
Chapman University, millerda@chapman.edu

Janice L. Krieger
Ohio State University

Jiangxiu Zhou
Pennsylvania State University

See next page for additional authors

Follow this and additional works at: http://digitalcommons.chapman.edu/comm_articles

Part of the Curriculum and Social Inquiry Commons, Education Economics Commons, Inequality and Stratification Commons, Medicine and Health Commons, Rural Sociology Commons, and the Substance Abuse and Addiction Commons

Recommended Citation

Random Assignment of Schools to Groups in the Drug Resistance Strategies Rural Project: Some New Methodological Twists

Authors

John W. Graham, Jonathan Pettigrew, Michelle Miller-Day, Janice L. Krieger, Jiangxiu Zhou, and Michael L. Hecht

Random Assignment of Schools to Groups in the Drug Resistance Strategies Rural Project: Some New Methodological Twists

John W. Graham,
Department of Biobehavioral Health, The Pennsylvania State University, University Park, PA, USA

Prevention Research Center, The Pennsylvania State University, University Park, PA, USA

Department of Biobehavioral Health, The Pennsylvania State University, 219 BBH Bldg, University Park, PA 16802, USA

Jonathan Pettigrew,
Department of Communication Arts and Sciences, The Pennsylvania State University, University Park, PA, USA

Michelle Miller-Day,
Department of Communication Arts and Sciences, The Pennsylvania State University, University Park, PA, USA

Janice L. Krieger,
School of Communication, The Ohio State University, Columbus, OH, USA

Jiangxiu Zhou, and
Department of Biobehavioral Health, The Pennsylvania State University, University Park, PA, USA

Michael L. Hecht
Prevention Research Center, The Pennsylvania State University, University Park, PA, USA

Department of Communication Arts and Sciences, The Pennsylvania State University, University Park, PA, USA

John W. Graham: jgraham@psu.edu

Abstract

Random assignment to groups is the foundation for scientifically rigorous clinical trials. But assignment is challenging in group randomized trials when only a few units (schools) are assigned to each condition. In the DRSR project, we assigned 39 rural Pennsylvania and Ohio schools to three conditions (rural, classic, control). But even with 13 schools per condition, achieving pretest

equivalence on important variables is not guaranteed. We collected data on six important school-level variables: rurality, number of grades in the school, enrollment per grade, percent white, percent receiving free/assisted lunch, and test scores. Key to our procedure was the inclusion of school-level drug use data, available for a subset of the schools. Also, key was that we handled the partial data with modern missing data techniques. We chose to create one composite stratifying variable based on the seven school-level variables available. Principal components analysis with the seven variables yielded two factors, which were averaged to form the composite inflate-suppress (CIS) score which was the basis of stratification. The CIS score was broken into three strata within each state; schools were assigned at random to the three program conditions from within each stratum, within each state. Results showed that program group membership was unrelated to the CIS score, the two factors making up the CIS score, and the seven items making up the factors. Program group membership was not significantly related to pretest measures of drug use (alcohol, cigarettes, marijuana, chewing tobacco; smallest $p>.15$), thus verifying that pretest equivalence was achieved.

Keywords

Random assignment; Group randomized trial; Principal component analysis; Missing data

Random assignment to groups is the foundation for scientifically rigorous clinical trials. The theory behind random assignment depends on the "long run." That is, random assignment is likely to work to produce pretest equivalence in treatment and control groups only when a sufficiently large number of units are assigned to each group. This is especially key in group randomized trials in which whole schools, or other intact units, are assigned to conditions. In such situations, it is rare that more than a few schools can be assigned to each condition, which makes ensuring pretest equivalence even more important in order to test accurately for intervention effects. The purpose of this article is to report on the random assignment procedures used in a school-based intervention study. Our approach, which involved a relatively small number of units (schools), illustrates the utility of applying missing data procedures in order to incorporate state substance use monitoring data into the random assignment process.

The Drug Resistance Strategies Rural Project

The Drug Resistance Strategies Rural (DRSR) project was designed to study adaptation processes for the school-based substance abuse prevention program known as *keepin' it REAL* (*kiR*; Colby et al. 2013; Hecht et al. 2003; Hecht and Miller-Day 2007). We studied two approaches to curricular adaptation:

Designer Adaptation

The original developers adapted the "Classic" curriculum, which made use of materials originally developed for a largely Latino, urban population, resulting in the new, "Rural" version of *kiR*. New written and video materials involving rural PA students were developed so as to be more relevant for the target, largely Caucasian, population. The designer adaptation processes used in DRSR are described in Colby et al. (2013).

Implementer Adaptation

We examined how teachers in largely Caucasian, rural schools adapted both the "Rural" version and the original "Classic" version of *kiR*. Papers describing some types of adaptations and reasons for implementer adaptation of *kiR* are described in Miller-Day et al. (2013).

The two versions may both have an impact on substance use. Many prevention programs are not developed with cultural diversity in mind (Harthun et al. 2008; Hecht et al. 2003) and, so, may require adaptation for different cultural audiences. Based on the principle of cultural grounding (Hecht and Krieger 2006), the designer-adapted version is hypothesized to reduce use because of its cultural alignment with the target population. That is, the designer-adapted version is expected to be more effective because it uses materials from within the rural, adolescent culture whereas the implementer adapted version imports material from a different culture (i.e., urban, Latino/a). Adaptations to the curriculum by the implementers may also affect program outcomes. Other studies report that teachers within classrooms of ethnic minority students were inclined to adapt prevention curricula to make lessons more culturally appropriate for their students (Ringwalt et al. 2004), and thus we hypothesized that rural teachers also may effectively adapt the classic curriculum to fit their students' culture. As data become available the effects of these two versions will be examined.

In the DRSR project, we originally recruited and assigned 41 schools to three experimental conditions. But even with 13–14 schools per condition, it was still essential to maximize the chance that random assignment would achieve pretest equivalence on important study variables, including the main dependent variable, student drug use.

One assignment strategy used under these circumstances is matching. With this strategy, each treatment school is matched with a control school such that the two schools share important pre-existing characteristics. Unfortunately this strategy has drawbacks. Matching does produce good comparability on all variables involved in the matching, but there is no guarantee that matching will produce pretest equivalence on variables not involved in matching. Also, there is a certain, perhaps well-deserved, stigma attached with non-random assignment procedures. Finally, program effects analyses with matched samples are not always straightforward.

A second strategy under these circumstances has been to use a stratified random assignment procedure. Stratification has advantages over matching because it is still a random assignment procedure. However, when relatively few intact units are to be assigned to each experimental condition, only a small number of stratifying variables can be used.

Finally, a stratifying procedure was suggested by Graham et al. (1984; also see Dent et al. 1993), in which numerous school-level variables can be taken into account and combined in a way that allows stratification to be handled with just a single stratifying variable. The procedure has been used successfully in several large, school-based prevention studies (e.g., Caldwell et al. 2012; Dent et al. 1993; Graham et al. 1984; Hansen and Graham 1991; Hecht et al. 2003). We used a variant of this procedure in the current project.

In school-based prevention studies such as this, it is typically possible, prior to random assignment, to obtain archival data on several relevant school-level variables. Graham et al. (1984) obtained school-level data for test scores, ethnic make-up (percent Anglo, Black, Hispanic, Asian-Pacific Islander), enrollment, SES (title I index), absences and mobility, crime incidents divided by school size, and percent nonfluent English speakers. These researchers also were able to judge each school's likely cooperation based on an overall rating by two school-district researchers with extensive experience with the various schools, and on prior school cooperation on recent nutrition and smoking studies carried out in the district. Dent et al. (1993) made use of a similar set of school-level variables, including school enrollment, number of sixth-grade classes, number of grades in the school, ethnic composition (percent white, black, Hispanic, and other), percent of students with English as a second language, SES (e.g., via percent of students receiving Aid to Families with Dependent Children), and test scores (reading, writing, and math).

One thing missing from the Graham et al. (1984) and Dent et al. (1993) studies were pre-assignment, school-level data on the main dependent variable: student drug use. As is often the case with this type of study, school recruitment and assignment to experimental conditions must occur in rapid succession, precluding the very desirable option of collecting pre-assignment data on the same schools that will be taking part in the study.

Assignment Procedure in the DRSR Project

Our goal for the current study was to follow the procedure outlined by Graham et al. (1984) to arrive at a single composite stratifying variable for randomization based on the school-level variables available. In the current study, we were also able to obtain measures of several important school-level variables relevant to study hypotheses: a rurality index, number of grades in the school, enrollment per grade, percent white, percent receiving free/assisted lunch, and test scores.

As noted above, it is usually not possible to incorporate school-level drug use data into decisions about random assignment because access to the subject population is typically limited until after random assignment is complete. Fortunately, the Pennsylvania Youth Survey (PAYS), which is sponsored and conducted biennially by the Pennsylvania Commission on Crime and Delinquency, had recently been conducted (2007). Thus, drug use data from a random sample of students were available for the majority of Pennsylvania schools involved in our study. Having drug use data for random assignment was highly desirable, to be sure. But in the past, having these data for only some of the schools to be randomly assigned would have been useless. However, modern missing data analysis techniques (multiple imputation and maximum likelihood) are available for addressing the missing school-level data (e.g., Graham 2009, 2012; Graham et al. 2003, 2012; Schafer and Graham 2002), and allowed us to make use of the partial data. The small number of schools involved in the assignment could have been a challenge, but Graham and Schafer (1999) have shown that multiple imputation performs very well even with small sample sizes. Still, 15 schools with drug use data is an extremely small number; thus, the issue of small sample size must be addressed.

In this manuscript, we present solutions to problems related to missing school-level drug use data as well as small sample sizes involved in group randomized trials, making a valuable contribution to the study of school-based interventions.

Method

Subject Population

The focus of the present study is on random assignment of the schools recruited for the DRSR project. Study participants were seventh grade students in the schools recruited for the DRSR study. The schools were recruited from largely rural parts of PA and OH. Based on the definitions provided by the National Center for Education Statistics (http://nces.ed.gov), 54 % of the recruited schools fell into the "rural" category; 29 % fell into the "town" category, and 17 % fell into the "suburb" category. Of the 41 schools initially recruited, the total enrollment median was 480 students. The median number of grades per school was six; and the median enrollment per grade was 104. The median percent of students described as white, non-Hispanic was 98 %; and the median percent eligible for free or reduced lunch was 32 %.

Archival, School-Level Measures

Form the National Center for Educational Statistics (NCES) Common Core of Data website, we were able obtain data for several school-level variables for both PA and OH schools.

Rurality Index (Rurality)—The NCES described 12 kinds of communities in four clusters: city, suburb, town, and rural. For our study, we used schools from the latter three clusters. Our rurality index was a nine-point scale ranging from "Rural Remote" (9) to "Suburb Large" (1).

School Configuration—In our data, schools fell into one of three basic configurations: grades 7–12 (including one 6–12, and one 6–11), 6–8 (including several 5–8 and some 7–8), and K-8 (including three preK-8). This latter configuration appeared in six OH schools only.

Number of Grades in the School (Numgrades)—Dent et al. (1993) used this variable in their analyses.

Enrollment per Grade (Npergrade)—This variable was total enrollment divided by number of grades in the school. Given the range of grades per school (two to nine), we felt that enrollment per grade was a more relevant indicator than total enrollment.

Percent White (Pctwhite)—Given that the overwhelming majority of students in participating schools were described as white, non-Hispanic (median=98 %), this was the only ethnicity-related variable used in our calculations.

Percent Receiving Free or Assisted Lunch (Pctlunch)—This variable, which is a commonly used indicator of SES, was calculated by summing two percentages: percent of total enrollment eligible for free lunch; percent of total enrollment eligible for reduced lunch.

Test Scores (Scores)—From the PA and OH Departments of Education web sites we were able to obtain information on recent math and reading test score performance for each of the schools. Although the tests were not the same in the two states, they were comparable. In PA schools, scores were obtained for whole schools; in OH schools, scores were obtained separately for sixth, seventh, and eighth grades. The scores used in OH schools were the average of available scores across these three grades.

School-Level Drug Use Data (Drugs)—The school-level drug use variable used in the imputation analysis (drugs) was the standardized average of eight individual drug use variables found in the PAYS data: lifetime and past-month use of alcohol, cigarettes, marijuana, and smokeless tobacco. A preliminary principal components analysis showed that all eight school-level variables loaded highly on a single factor. The Drugs variable was included, without modification, in the imputation model.

Missing Data Imputation

Additional Sample of PA Schools—As described above, school-level drug use data came from the PAYS study. These data were available for 15 of the 26 PA schools in our study. In order to improve the stability of the imputation model, we included in the imputation analysis 28 additional PA schools not included in the DRSR study, but for which the PAYS drug use data were available, as were data for all of the other school-level, archival data described above. This strategy is akin to what is known in the statistics literature as using an "informative prior" for imputation (e.g., Schafer 1997). The strategy makes sense only when it is reasonable to assume that the prior information comes from the same population as the main data. In this case, we judged this to be a very reasonable assumption. Fourteen of the additional 28 schools were selected from the pool of schools we had previously sought to recruit for the DRSR study, and 14 others were randomly selected from the remaining list of PAYS schools. All 28 schools selected here would have been eligible to be part of our study based on rurality criteria, and on the presence of seventh grade class in the school. By all accounts, this group of 28 schools was very similar the group of 26 PA schools taking part in the DRSR study; there were no significant differences on assignment variables between the 26 schools in the study and the 28 additional schools (analysis not shown). Table 1 summarizes the sample sizes for the different groupings of PA and OH schools with different patterns of observed and missing data for the school-level, archival data.

The Missing at Random (MAR) Assumption—It is commonly noted that modern missing data procedures (multiple imputation, maximum likelihood, and related procedures) assume that the missingness is MAR (e.g., Little and Rubin 2002; Schafer and Graham 2002). What is not often said, however, is that these procedures work well, and are far superior to old procedures (e.g., listwise deletion analysis), even when the MAR assumption is violated (Graham 2009; 2012). Some believe that checking for differences, in this case between schools with and without PAYS data, can help the researcher determine whether or not the MAR assumption holds. Although this type of test has intuitive appeal, it is not true that such tests can help the researcher know if the MAR assumption has been violated. Although tests do exist that can at least help in this regard, they apply to longitudinal data,

and not to the kind of cross-sectional data available in the present study (Graham 2009; 2012; also see Little 1994, p. 482). In any case, based on available information, we estimate that the marginal probability is 92 % that any PA school without PAYS data was never asked to take part in the PAYS study. In sum, we have every reason to believe that MAR holds reasonably well in the present study, and we argue that in this situation, the most important issue is whether pretest equivalence was achieved, and the degree to which missingness was MAR is of secondary importance.

The expanded school-level data set had N=26+28+15= 69 PA and OH schools. This number (69) compares favorably with the sample size (50) used in simulations showing that normal-model MI works well with small samples (Graham and Schafer 1999). With two exceptions (for the Percent White variable), the only missing values were for the school-level drug use variable as shown in Table 1. Following Graham (2012; also see Graham et al. 2012), we performed a single imputation from EM parameters using SAS (version 9.2) Proc MI (Proc MI adds random normal error to each imputed value) on the data set with N=69 cases (schools).

School Configuration Dummy Variable—Given the three grade configurations described above, two dummy codes were used to represent these variations. The variable S5678 was "1" if school type was "5678," otherwise it was "0." The variable SK8 was "1" if school type was "K-8," otherwise it was "0." School type "7–12" was "0" for both dummy variables. We also used a dummy variable to represent the state (PA vs. OH; PAOH).

Single Imputation of School-Level Variables—Ideally, all of the school-level variables, the raw, school-level, drug use variable from the PAYS data, and the dummy variables representing state (PAOH) and grade configuration (S5678 and SK8), would all be included in the imputation analysis. However, because all drug use data were missing for OH schools, including the PAOH dummy variable would produce pathological imputation (see Graham 2012; Schafer 1997). Similarly, because the K-8 grade pattern was found only in the OH schools, and because the OH schools had no drug use data, using the SK8 dummy variable in imputation would also produce pathological imputation. Thus, in this instance, the imputation should be conducted using the S5678 dummy variable, but omitting the SK8 and PAOH dummy variables.

School-Level Drug Use to be Used in Analyses—The drugs variable described above was a reasonable estimate of drug use in each of the schools for which data were available. However, the three school types represent different average ages of the students. Assuming similar grade sizes, the average student in grade 6–8 schools, for example, would be in grade 7; the average student in grade K-8 schools would be in grade 4; and the average student in grade 7–12 schools would be in grade 9.5.

These differences in grade would be expected to lead to differences in drug use among schools for normative, age-related reasons, and not for reasons that might be associated to the degree to which schools were at risk. In order to control for these artifactual differences, we regressed the drug use score on the two dummy variables representing school grade

configuration. The variable used in subsequent analyses was the residual from that analysis (Rdrugs), that is, drug use controlling for school configuration.

Assignment Procedure

Our goal was to follow the procedure outlined by Graham et al. (1984) to arrive at a single composite stratifying variable for randomization based on the school-level variables available. To begin, seven key variables from the imputed data set were subjected to principal components analysis. The promax rotated factor pattern for the seven, key, school-level variables is shown in Table 2.

Interpretation of the Factors—The primary purpose of the principal components analysis in this context is to find combinations of variables that are closely related in the study sample. It is a bonus if the combinations of variables loading on each factor have common substantive meaning. In this instance, the first factor (component) was dominated by the number of grades in the school, percent white, enrollment per grade, and rurality. This factor could be thought of as a "rurality" factor. Schools in more rural areas tend to cover more grades, have smaller numbers of students per grade, and tend to have a larger proportion of white students.

The second factor appears to be somewhat more broadly defined. Schools with higher drug use also tend to have lower test scores. In this case, higher drug use and lower scores are also associated with a higher proportion of students receiving free or assisted lunch (we also explored performing the imputation and principal components analyses separately within PA and OH schools; despite the lack of the Rdrugs variable in the OH schools, the results of the principal components analysis within each state were very similar to what appears in Table 2).

The Composite Inflate-Suppress (CIS) Continuum—In the next step of the assignment procedure, Graham et al. (1984) suggested thinking about the effects on internal validity of pretest differences on these two factors. Often it is easy to see these effects for one or more of the factors. Suppose, for example that schools in the present study that were high on factor 2 (higher drug use, lower test scores) were assigned to the program group, and schools low on factor 2 (lower drug use, higher test scores) were assigned to the control group. In this instance, it is rather clear that this kind of pretest difference would tend to suppress true program effects (program would look less effective than it really was). On the other hand, pretest differences on other factors may be less clearly interpreted. In our study, for example, it is not as obvious whether pretest differences on Factor 1 would tend to suppress or inflate apparent program effects.

Combining Factors: The CIS Score—In the Graham et al. (1984) procedure, all factors were recoded so that all were in the same direction in the single CIS continuum, such that, for example, high values for all factors were associated with the "inflate" end of the continuum. However, even in the present study, where only one of the factors could be placed clearly on that continuum, it makes sense to combine the factors to obtain the CIS Score in order to create a single stratification variable incorporating all relevant information.

In this instance, because the two factors were observed to be positively correlated ($r=.22$; see Table 2), it made sense to combine the two factors in their original form. As recommended in Graham et al. (1984), we generated factor scores for the two factors, and simply averaged them to create the set of CIS scores.

Stratified Random Assignment to Groups—The CIS score is a single variable that is a good basis for stratified random assignment to groups. Factor scores, on which the CIS score was based, are standardized variables (mean=0; standard deviation=1). Thus, the overall CIS score retains the mean of 0, but the standard deviation for the average of the two factor scores will always be somewhat smaller than 1, typically around .75, depending on the correlation between the factor scores.

The CIS score took on a unique value for the 26 PA and 15 OH schools that were part of the DRSR project. We broke the CIS score into three strata within each state. Within PA schools the eight schools with the lowest CIS scores (highest test scores, least drug use, fewest grade levels, least white, least rural) were in stratum 0, the nine schools with intermediate CIS scores were in stratum 1, and the nine schools with the highest CIS scores (lowest test scores, most drug use, most grade levels, most white, most rural) were in stratum 2. Within the OH schools, the three groups had the same meaning, but to help with the assignment to three conditions, there were three schools in stratum 0, six schools in stratum 1, and six schools in stratum 2.

We then assigned schools at random to the three program conditions (rural, classic, control) from within each stratum, within each state. Of the 26 PA schools recruited, nine were assigned randomly to the rural and control conditions, and eight were assigned to the classic condition. Of the 15 OH schools originally recruited, five were assigned randomly to each of the three conditions (rural, classic, and control).

Preliminary Tests of Equivalence

First, we verified that the assignment procedure worked to provide group equivalence on the school-level variables relating to CIS score. We conducted a simple ANOVA (SAS Proc GLM), with Program (Rural, Classic, Control) listed as a class variable. Table 3 presents the results for the overall F test. As expected, program group membership was unrelated to the CIS score, the two factors making up the CIS score, and the seven items making up the factors. Note that these tests were conducted on the imputed data set. The only variable truly affected by the missing data was the rdrugs variable; what is shown here represents a conservative estimate of the program effect on this variable.

"Random" Assignment to Groups in the Real World: Post-Assignment

Adjustments Initially, 41 schools were assigned to the three conditions using a randomizing procedure described above. The assignment procedure involved developing three strata, within each state, based on the several school-level variables described above. Schools within each state were then randomly assigned to the three conditions from within each of the strata. As described above, the initial random assignment was successful in creating three conditions that were not significantly different on any of the school-level factors. It

was also successful in generating nearly equal numbers of schools, from each of the strata, within each state. Table 4 shows the relevant assignment numbers based on the original assignment.

After random assignment was complete, and after several pre-intervention steps had already been taken that would preclude redoing the random assignment, six schools withdrew from the study (three from OH, three from PA). Schools withdrawing after assignment is, unfortunately, all too common in school-based research. It happened that four of the withdrawn schools were from condition 1 (DRS classic), and two were from condition 2 (DRS rural). The numbers of schools withdrawing from each stratum/condition combination and from each condition, overall, are represented using negative superscripts in Table 4.

Four replacement schools were found, all from OH, resulting in a final total of 39 schools. Within OH, two of the lost schools were from stratum 0 (highest test scores, lowest numbers of free lunch, and lowest rurality). One of these schools was in condition 1; one was in condition 2. As it happened, two of the replacement schools were also in this stratum. It seemed reasonable, then, to reassign these new schools to the two conditions. Thus, one of these was randomly assigned to be in condition 1 (DRS classic); the other was assigned to be in condition 2 (DRS rural).

The remaining lost school in OH was in stratum 1 (intermediate test scores, school lunches, and rurality). This school had been assigned to condition 2 (DRS Rural). Only one of the replacement schools was in this stratum, so that school was assigned to condition 2. The final replacement school, which was in stratum 2 (lowest scores, highest school lunches, highest rurality), was assigned to condition 1 because a total of four schools had been lost from condition 1 and only one replacement school had been assigned to that condition. Positive superscripts are used in Table 4 to show the stratum/combinations of the replacement schools.

Tests of Equivalence on Pretest Data

Finally, after making post-assignment adjustments, we conducted analyses on the 39 schools to determine if pretest equivalence was achieved on the main drug use-dependent variables under study. In this case, we performed a mixed effects analysis using SAS Proc Mixed. The dependent variables for these analyses (summarized in Table 5) were the lifetime use, use in the previous 30 days of alcohol, cigarettes, marijuana, and chewing tobacco. For example, the lifetime alcohol item was, "How many drinks of alcohol have you had in your entire life?" (responses ranged from 1="none" to 10="More than 100 drinks"); the 30-day smoking item was, "How many cigarettes have you smoked in the past 30 days? " (responses ranged from 1="none" to 8="More than two packs of cigarettes"). The mixed effects analysis treated intercept as a random effect, thereby controlling for the intraclass correlation in the analyses. We tested two orthogonal effects in these analyses: (a) any program vs control (rural+classic vs control), and (b) rural vs classic. The results of these tests are summarized in Table 5.

Program group membership was not significantly related to pretest measures of any of the main DVs (use of alcohol, cigarettes, marijuana, chewing tobacco; smallest $p=.15$), thus verifying that pretest equivalence was achieved (see Table 5).

Discussion

The random assignment procedure used in the present study followed the procedure outlined by Graham et al. (1984). As with the original version, our procedure combined several school-level variables that could, if not properly controlled, have an undesirable effect on the internal validity of the study. New to this study was the addition of relevant archival, school-level, drug use data for many (but not all) of the schools taking part in our study. Also, new to this study was the use of a modern missing data imputation procedure. Because of this, we were able to make use of the partial drug use data, and to incorporate these highly relevant data into our assignment procedure. As with the original study, assignment produced experimental and control conditions for which no significant differences existed for the CIS score, for the two factors making up that score, or for any of the seven school-level variables used in the procedure.

To be sure, it is never desirable to lose participating schools after random assignment is complete. However, losing participating schools at various points along the way is a common reality of doing research in the real world. Although the ad hoc replacement of schools does force a deviation from true random assignment, the process and impact of such replacement of schools was relatively painless in our study because we were able to make use of a simple extension of our assignment procedure. It was an easy matter to place possible replacement schools into the three strata of the CIS score, and to assign new schools to conditions in a way that would minimize pretest differences between groups. It is a testimony to these efforts that our final assignment produced groups that showed no statistically significant differences on any of the drug use variables that will serve as the primary dependent variables in our study.

The procedure described in this article should prove an asset to those conducting group randomized trials that are common in the prevention literature. This study offers further evidence in support of the random assignment procedure initially described by Graham et al. (1984). It also extends that earlier work by offering a useful missing data analysis alternative when some of the key school-level data are missing at the time the randomization is conducted. Limitations to this method as well as alternative assignment approaches vis-à-vis the procedure used in this study are discussed below.

Limitations to Assignment Procedure

Statistical Power—One important consideration in group randomized trials is pretest equality, something the procedure outlined in this study addresses; however, equally important is maintaining sufficient power to detect intervention effects. In the case of the DRSR project, as is typical of many school-based intervention studies, power issues were dealt with in our original proposal to the funding agency. At that time, the proposed number of schools (39) was deemed to have sufficient power to detect effects of the magnitude

expected. At the assignment stage of the project, there was no new information available to update our power estimates.

An additional power consideration is the number of strata used in assignment. In our view, less stratifying is best in contexts where few level 2 units (schools) are available for randomization. Statisticians recommend that the analysis model should take the stratification into account (e.g., see Kahan and Morris 2012; Matts and Lachin 1988). When there are few stratifying variables, the proper analysis is straightforward. However, as the number of stratifying variables (and levels) increases, making use of the proper analysis becomes more of difficult. A major concern is that including the stratifying variables as covariates in the analysis model takes degrees of freedom away from the already limited denominator degrees of freedom, which can reduce statistical power. This is especially true when there are very few level 2 units assigned to each treatment condition. In our procedure, with relatively few strata used in the assignment, there is nothing that would materially affect the power to detect treatment effects provided that the later program effects analyses take stratification appropriately into account. These points will typically be true with any project considering using this assignment procedure.

Inflate-Suppress Continuum—Another issue involved in using the assignment procedure employed in this study is how to assign factors to the inflate-suppress continuum. In our study, one factor clearly would suppress true program effects while the effect of the other factor was not obvious. In other studies, it can also be difficult to judge with any accuracy whether a particular factor will have an inflating or suppressing effect on the program. Also, although we were not able to think of any examples, it remains a possibility that two factors (or items within factors, for that matter) could be positively correlated and still have one be inflating and the other suppressing. However, it may well be that neither of these issues is a major concern. Although more research should be done to bear this out fully, preliminary simulation work with our current data showed that the group equivalence results on the seven school-level variables were essentially the same whether one factor was added to the other, or one factor was subtracted from the other. That is, at least preliminary evidence suggests that the benefits of averaging are found regardless of how the factors are scaled.

Alternate Assignment Approaches to Assignment

Available methods for addressing pretest equivalence of conditions in randomized trials have ranged from simple random assignment of subjects (units), at one extreme, to individual subject (unit) matching at the other extreme. The advantage of simple random assignment is that all subject variables, measured and unmeasured, will be equivalent (with a known probability) in the treatment and control groups. A further advantage of using simple random assignment is that the analysis of the treatment effects is straightforward (e.g., ANOVA, ANCOVA, or regression analysis and similar methods).

On the other hand, the advantage of matching is that measured variables involved in matching are guaranteed to be equivalent at pretest in treatment and control groups. However, with matching there is no guarantee that unmeasured subject variables will also be

equivalent in treatment and control groups. Further, because one subject (unit) in the treatment group is yoked with one subject (unit) in the control group, different statistical analyses are required.

Randomizing procedures that move away from simple random assignment, most notably stratified random assignment, retains much of the benefit of simple random assignment. With stratified random assignment, pretest equivalence is more likely to occur on variables involved in the stratification, even with somewhat smaller sample sizes, but pretest equivalence is still guaranteed for unmeasured subject variables with a certain probability. The drawback is that treatment effect analyses must be modified somewhat to take the stratifying variables into account. When just one variable is used for stratification, the modified analysis, including the stratifying variable(s) as covariates in the regression model, is a simple departure from the initial regression analysis. In fact, with just a single stratifying variable, there could well be a net gain in statistical power. However, more substantial departures from simple random assignment may not have such desirable qualities. Having two or more stratifying variables, for example, could produce a net loss in statistical power over simple random assignment, mainly due to loss of denominator degrees of freedom in the analysis.

Other approaches (e.g., see Xu and Kalbfleisch 2010) appear to be between simple random assignment and subject matching. With the Xu and Kalbfleisch approach (which they refer to as the BMW approach), simple random assignment of subjects to conditions is repeated several times (e.g., 10–20 times), and the optimal solution is obtained with respect to propensity score distances. Our opinion is that this procedure, and others like it, will, like matching, produce better pretest equivalence on measured variables involved in the assignment procedure. However, because the procedure makes use of matching, the guarantees of simple random assignment for equivalence on all unmeasured variables will necessarily be weakened. Further, because of the optimization approach taken in selecting the particular assignment used, it is not clear what analysis adjustments would need to be taken in order to obtain proper statistical conclusions.

We conclude that there are essentially three types of assignment procedures. One is simple random assignment including stratified random assignment with very few stratifying variables; a second is subject matching; and the third is a hybrid approach, such as that suggested by Xu and Kalbfleisch (2010) that seem to take the middle ground between simple random assignment and complete matching.

Although all three of these assignment approaches have value in different research contexts, we believe that the simple random assignment/limited stratified random assignment will work best in school-based prevention research. Generally in the school-based context there are relatively little data available prior to assignment, and data that are available are typically school-level data from previous years. Indeed, practical issues typically determine what cluster-level variables are available for use in assignment, and a researcher is often at the mercy of what archival data can be located. Also important is that available archival data are at best proxies for the possible confounds a researcher would ideally measure, and it is typically not feasible to collect new data prior to randomization. Despite these limitations,

even if the procedure outlined here moves the study toward pretest equivalence, it has done its job. That is, the goal of the procedure should not be to achieve exact pretest equivalence, but rather to help the researcher avoid the worst case scenario with respect to pretest equivalence. Thus, the procedure described in this study, we believe, is best for the school-based research context.

Other approaches (i.e., hybrid, matching) may be more applicable in research contexts with different constraints and opportunities. Moving away from random assignment, which ensures pretest equivalence on unmeasured variables, only seems advisable under certain conditions. Research contexts where better measures of potential confounds are available at the time of randomization, for example, might usefully apply a hybrid approach. More research, including simulation studies, is needed to further investigate these hypotheses and could demonstrate under what conditions each assignment procedure is optimal.

Acknowledgments

References

Caldwell LL, Smith EA, Collins LM, Graham JW, Lai M, Wegner L, Jacobs J. Translational research in South Africa: Evaluating implementation quality using a factorial design. Children and Youth Care Forum. 2012; 41:119–136.

Colby M, Hecht ML, Miller-Day M, Krieger JR, Syvertsen AK, Graham JW, Pettigrew J. Adapting school-based substance use prevention curriculum through cultural grounding: A review and exemplar of adaptation processes for rural schools. American Journal of Community Psychology. 2013; 51:190–205. [PubMed: 22961604]

Dent CW, Sussman S, Flay BR. The use of archival data to select and assign schools in a drug prevention trial. Evaluation Review. 1993; 17:159–181.

Graham JW. Missing data analysis: Making it work in the real world. Annual Review of Psychology. 2009; 60:549–576.

Graham, JW. Missing data: Analysis and design. New York: Springer; 2012.

Graham, JW.; Cumsille, PE.; Elek-Fisk, E. Methods for handling missing data. In: Schinka, JA.; Velicer, WF., editors. Research methods in psychology. New York: Wiley; 2003. p. 87-114.

Graham, JW.; Cumsille, PE.; Shevock, AE. Methods for handling missing data. In: Schinka, JA.; Velicer, WF., editors. Research methods in psychology. 2nd ed.. New York: Wiley; 2012. p. 109-141.

Graham JW, Flay BR, Johnson CA, Hansen WB, Collins LM. Group comparability: A multiattribute utility measurement approach to the use of random assignment with small numbers of aggregated units. Evaluation Review. 1984; 8:247–260.

Graham, JW.; Schafer, JL. On the performance of multiple imputation for multivariate data with small sample size. In: Hoyle, R., editor. Statistical strategies for small sample research. Thousand Oaks: Sage; 1999. p. 1-29.

Hansen WB, Graham JW. Preventing alcohol, marijuana, and cigarette use among adolescents: Peer pressure resistance training versus establishing conservative norms. Preventive Medicine. 1991; 20:414–430. [PubMed: 1862062]

Harthun ML, Dustman PA, Reeves LJ, Hecht ML, Marsiglia FF. Culture in the classroom: Developing teacher proficiency in delivering a culturally-grounded prevention curriculum. The Journal of Primary Prevention. 2008; 29:435–454. [PubMed: 18807191]

Hecht ML, Krieger JK. The principle of cultural grounding in school-based substance use prevention: The Drug Resistance Strategies Project. Journal of Language and Social Psychology. 2006; 25:301–319.

Hecht ML, Marsiglia FF, Elek E, Wagstaff DA, Kulis S, Dustman P, Miller-Day M. Culturally grounded substance abuse prevention: An evaluation of the *keepin' it REAL* curriculum. Prevention Science. 2003; 4:233–248. [PubMed: 14598996]

Hecht ML, Miller-Day M. The drug resistance strategies project as translational research. Journal of Applied Communication Research. 2007; 35:343–349.

Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. Statistics in Medicine. 2012; 31:328–340. [PubMed: 22139891]

Little RJA. A class of pattern-mixture models for normal incomplete data. Biometrika. 1994; 81:471–483.

Little, RJA.; Rubin, DB. Statistical analysis with missing data. 2nd ed.. Hoboken: Wiley; 2002.

Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. Controlled Clinical Trials. 1988; 9:327–344. [PubMed: 3203524]

Miller-Day M, Pettigrew J, Hecht MLM, Shin Y, Graham J, Krieger J. How prevention curricula are taught under real-world conditions: Types of and reasons for teacher adaptations. Health Education. 2013; 113(4) in press.

Ringwalt CL, Vincus A, Ennett S, Johnson R, Rohrbach LA. Reasons for teachers' adaptation of substance use prevention curricula in schools with non-white student populations. Prevention Science. 2004; 5:61–67. [PubMed: 15058914]

Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychological Methods. 2002; 7:147–177. [PubMed: 12090408]

Schafer, JL. Analysis of incomplete multivariate data. New York: Chapman and Hall; 1997.

Xu Z, Kalbfleisch JD. Propensity score matching in randomized clinical trials. Biometrics. 2010; 66:813–823. [PubMed: 19995353]

Table 1

Patterns of observed and missing data for four categories of PA and OH schools

	NCES data	PA or OH Department of Education data	PAYS drug use data	Number of schools
PA school; DRS project, with drug use data	1	1	1	15
PA school; DRS project, no drug use data	1	1	0	11
PA school; not in DRS project	1	1	1	28
OH school	1	1	0	15

Table 2

Promax rotated factor pattern (standardized regression coefficients) from principal components analysis

	Factor 1	Factor 2
Numgrades	.85	−.02
Pctwhite	.75	−.19
Npergrade	−.71	−.25
Rurality	.67	.06
Rdrugs2 (resid)	−.18	.83
Scores	−.11	−.78
Pctlunch	.08	.68

Note: The correlation between the two factors was $r=.22$

Table 3

Significance tests for group equivalence on CIS-related variables

Variable	$F(2, 38)$	p
CIS	0.25	.78
Factor1	1.05	.36
Factor2	0.08	.93
Rurality	0.45	.64
Npergrade	1.62	.21
Numgrades	0.41	.67
Pctwhite	1.60	.21
Pctlunch	0.10	.90
Scores	0.22	.81
Rdrugs	0.80	.46

Note: Post hoc tests with the Duncan test showed no significant differences (p<.10) for any individual comparisons.

Table 4

Original assignment numbers

	Pennsylvania condition			Ohio condition		
	0	**1**	**2**	**0**	**1**	**2**
Stratum 0 (highest test scores)	3	2^{-2}	3	1	1^{-1+1}	1^{-1+1}
Stratum 1 (intermediate test scores)	3	3	3	2	2	2^{-1+1}
Stratum 2 (lowest test scores)	3	3^{-1}	3	2	2^{+1}	2
Overall	9	8^{-3}	9	5	5^{-1+2}	5^{-2+2}

Note: A negative superscript ($^{-1}$) indicates that a school in this stratum/condition combination withdrew from the study. A positive superscript ($^{+1}$) indicates a replacement school was in this stratum/condition combination. Test scores refers to academic achievement. Test scores were inversely related to substance use so that stratum 2 had the lowest test scores and highest drug use. Condition was assigned as 0=control; 1=classic; 2=rural

Table 5

Tests of program group differences on pretest measures of main drug use variables

DV	Effect			
	Any program vs Control		Rural vs Classic	
	F	p	F	p
Alcohol				
Lifetime alcohol use	1.03	.32	0.40	.53
Alcohol use past 30 days	0.09	.76	0.04	.84
Cigarette smoking				
Lifetime cigarette smoking	+2.15	.15	1.14	.29
Cigarette smoking past 30 days	+2.17	.15	0.25	.62
Marijuana				
Lifetime marijuana use	+1.43	.24	0.18	.67
Marijuana use past 30 days	0.26	.62	0.20	.66
Chewing tobacco				
Lifetime chewing tobacco use	0.33	.57	1.26	.27
Chewing tobacco use past 30 days	0.07	.79	1.10	.30

Note: For all results shown, numerator $df=1$; denominator $df=36$. Signs shown for F values refer to the sign of the effect. + means that the program groups had higher means than did the control group for that effect. Signs are shown only when $p<.25$