Prev Sci (2012) 13:300–313
DOI 10.1007/s11121-011-0265-y

Person Mobility in the Design and Analysis of Cluster-Randomized Cohort Prevention Trials

Sam Vuchinich · Brian R. Flay · Lawrence Aber · Leonard Bickman

Published online: 17 January 2012

Abstract Person mobility is an inescapable fact of life for most cluster-randomized (e.g., schools, hospitals, clinic, cities, state) cohort prevention trials. Mobility rates are an important substantive consideration in estimating the effects of an intervention. In cluster-randomized trials, mobility rates are often correlated with ethnicity, poverty and other variables associated with disparity. This raises the possibility that estimated intervention effects may generalize to only the least mobile segments of a population and, thus, create a threat to external validity. Such mobility can also create threats to the internal validity of conclusions from randomized trials. Researchers must decide how to deal with persons who leave study clusters during a trial (dropouts), persons and clusters that do not comply with an assigned intervention, and persons who enter clusters during a trial (late entrants), in addition to the persons who remain for the duration of a trial (stayers). Statistical techniques alone cannot solve the key issues of internal and external validity raised by the phenomenon of person mobility. This commentary presents a systematic, Campbellian-type analysis of person mobility in cluster–randomized cohort prevention trials. It describes four approaches for dealing with dropouts, late entrants and stayers with respect to data collection, analysis and generalizability. The questions at issue are: 1) From whom should data be collected at each wave of data collection? 2) Which cases should be included in the analyses of an intervention effect? and 3) To what populations can trial results be generalized? The conclusions lead to recommendations for the design and analysis of future cluster-randomized cohort prevention trials.

Keywords Mobility · Cluster-randomized trials · Cohort prevention trials · Validity · Generalizability

S. Vuchinich (✉)
School of Social and Behavioral Health Sciences,
Oregon State University,
314 Milam Hall,
Corvallis, OR 97331, USA
e-mail: vuchinis@oregonstate.edu

B. R. Flay
School of Social and Behavioral Health Sciences,
Oregon State University,
321 Waldo Hall,
Corvallis, OR 97331, USA
e-mail: brian.flay@oregonstate.edu

L. Aber
Steinhardt School of Culture, Education, and Human Development, New York University,
Kimball, 417E, 246 Greene St,
New York, NY, USA
e-mail: lawrence.aber@nyu.edu

L. Bickman
Department of Psychology and Human Development,
Vanderbilt University,
303 Peabody Administration Building,
Nashville, TN, USA
email: leonard.bickman@vanderbilt.edu

The mobility of persons into and out of clusters such as schools, neighborhoods, cities, or states is frequent in the US and many cultures. 12% of the total US population and 24% of those below the poverty line moved during 2008 (US Census Bureau 2009). The most important consequences of such mobility for prevention science concern the determination of whether or not preventive interventions actually have a preventive effect. Most well known in this regard are threats to valid causal inference in randomized trials where person mobility creates attrition from the trial (e.g., Barnard et al. 2003; Shadish et al. 2002). Differential attrition can introduce differences between intervention and control conditions on

measured and unmeasured variables that were not present at baseline assessment due to random assignment. In such trials, at endpoint, it becomes a challenge to determine whether observed outcome differences between conditions were due to the intervention or to differential attrition.

Person mobility can also create forms of noncompliance, where persons or clusters assigned to receive an intervention actually receive only part of it (e.g., Frangakis and Rubin 2002; Little et al. 2009). Indeed there are other forms of noncompliance such as when an intervention is not delivered or received as planned for any reason. In this context, any differences observed at trial endpoint may be due to any of three processes or some combination of them: 1) mobility processes that determine who is still in the trial at endpoint, 2) the extent of compliance with the intervention protocol, and 3) the actual effects of the intervention on those who complied with it. Unfortunately, these processes cannot be easily separated after trial data are collected (e.g., Bloom 2005; Frangakis and Rubin 2002). Estimates of intervention effects may be biased because any of these processes could produce observed differences at endpoint. Disentangling these processes has been widely recognized as a fundamental problem for drawing accurate inferences from preventive trials. We address this issue from the standpoint of validity and design decisions that can provide the most accurate estimate of intervention effects when mobility is present. Two primary strategies have emerged in the prevention sciences as methods to address mobility.

One strategy is to use research protocols that emphasize retention of persons in the trial to reduce mobility, attrition and noncompliance by design (Prinz et al. 2001; Ribisl et al. 1996; Shadish et al. 2002). These methods include sampling from populations that are known to have low mobility (as in the Dunedin Longitudinal Study; Danese et al. 2008), maintaining regular positive contact (mail, telephone, internet) with sample members, getting phone and address information for relatives likely to know future whereabouts (as in the Framingham Heart Study; e.g., Christakis and Fowler 2007), and using internet or private investigation services, monetary or other incentives for participation. Such methods can be especially effective in reducing mobility, and compliance problems in small-sample, short-term biological trials in populations with low mobility (e.g., Clark et al. 2009). However, reducing mobility and related problems by selective sampling can create limited external validity and the trial results may only apply to a limited subsample of the target population—those who remain in the same cluster and comply with the intervention.

The impact of person mobility on effect estimates is greater in trials that require more subjects, take more time to deliver an adequate dosage of the intervention, such as early interventions to prevent long-term outcomes or mass media campaigns (Flay and Cook 1981), and that target populations that are more mobile or at-risk (Kellam et al. 2008; Liu et al. 2009). In such cases societal disparities may also affect retention in the sample, as when minorities are more likely to drop out of a randomized trial (Osann et al. 2011). In trials that take more time to produce effects, protocol costs for tracking drop-outs can be prohibitive. Even when considerable resources are applied in at-risk samples, attrition in the range of 20% or more still occurs (Brown et al. 2008; DeGarmo et al. 2009; Dodge et al. 2008). Strategies for retention in randomized cohort prevention trials will continue to be useful in obtaining accurate estimates of intervention effects. But the attrition rates in recent preventive trials (e.g., Brown et al. 2008) suggest that the best of retention tactics yet available will not eliminate attrition as a problem in estimating the effects of randomized trials in some at-risk populations.

The second prevalent strategy for dealing with attrition and noncompliance caused by mobility is to use statistical techniques that make estimates about what differences between intervention and control groups *would have been observed* had there been no attrition or noncompliance created by person mobility or other factors. A wide variety of techniques have been presented to address the problem from a statistical perspective. These include multiple imputation (e.g., Collins et al. 2001; Graham 2009; Rubin 1987; Schafer and Graham 2002), potential outcome models (e.g., Frangakis and Rubin 2002; Jo et al. 2010; Stuart et al. 2008), maximum likelihood estimation (e.g., Little and Rubin 2002), growth mixture models (e.g., Muthen et al. 2002), propensity score methods (Jo and Stuart 2009) and others (e.g., Fitzmaurice and Laird 2000; Goldstein et al. 2007; Jo et al. 2008). These methods involve a variety of assumptions (Frangakis and Rubin 2002; Jo et al. 2010), or the availability of informative auxiliary information that can accurately predict attrition, noncompliance and the primary outcome measure (Baker et al. 2006; Jo and Muthen 2003; Leon et al. 2007; Rubin 2006).

These methods represent important advances in how to estimate intervention effects in the presence of attrition and noncompliance produced by person mobility and other processes. However, retention tactics in protocols and statistical techniques must be supplemented with other considerations when estimating intervention effects in the presence of person mobility. This commentary presents a third strategy that can complement the other two—systematic consideration of internal and external validity (Shadish et al. 2002) based on a cross-classification of whether the trial has a focus on person or cluster, and on whether it uses an intent-to-treat (ITT) or noncompliance analysis. The ITT analysis assesses all randomized units (persons or clusters) at endpoint regardless of whether they complied with the intervention or not. The noncompliance analysis takes into account whether they complied or not.

Ideally, randomized cohort prevention trials have no mobility, attrition or noncompliance. For some prevention trials these conditions can be met (e.g., Clark et al. 2009) and have been classified by some as "pharmaceutical randomized clinical trials" (Brown et al. 2008). In such trials the ITT estimation of preventive effects is straightforward, well known and unbiased. However, this requires that the randomization was successful in creating condition (intervention and control) balance on measured potential confounders, and that the untestable assumption that randomization also created condition balance on unmeasured potential confounders. But when there is attrition or noncompliance due to person mobility or other processes, estimation of trial effects are more complicated.

As prevention science has progressed, interventions have been developed to prevent important negative outcomes in some populations where person mobility, attrition and noncompliance are unavoidable. An ITT analysis is always important from a practical standpoint because it yields an overall unbiased estimate of the effect of *assignment to receive* an intervention with the fewest assumptions. But when there is attrition and noncompliance, that intervention effect estimate may be a composite of: 1) the actual effect of the intervention on those who complied with it, 2) the effect on those who didn't comply even though they were assigned to receive it, and 3) any effect on those who dropped out of the trial before an adequate dosage was received. Though the ITT analysis is always useful, progress toward improving interventions based on the results of previous trials requires more specific information about response to an intervention. For example, was the intervention effective on those who complied with it? If so, then improvements in preventive effect could be obtained by *increasing compliance*, aside from elements of the intervention itself. And, furthermore, was there was an intervention effect on those who dropped out of the trial before an adequate dosage was received? This information could be useful in determining how much of the intervention is needed to yield a beneficial effect.

Outside of the realm of statistical considerations, distinctions among research designs for preventive trials have been introduced that implicitly acknowledge variations in attrition and compliance associated with mobility. These variations include the distinction between "efficacy" trials, those done under ideal conditions with good compliance and low attrition, and "effectiveness" trials, those done under real-world conditions with more attrition and noncompliance (Flay 1986).

Also influential for attrition and noncompliance issues has been the emergence of cluster-randomized (or group-randomized) cohort trials (e.g., Bloom 2005; Brown et al. 2008; Cornfield 1978; Donner and Klar 2000; Murray 1998; Sherwin 1978), in which random assignment to intervention or control group is done based on clusters or groups (e.g., clinics, schools, hospitals, counties, states, physicians) rather than individual persons. In such trials, compliance and attrition can be more heavily regulated by professionals (e.g., physicians, teachers) and institutions (e.g., clinics, schools, hospitals) rather than relying only on individual persons to adhere to an intervention. In addition, such designs limit the possibility of contamination of conditions because some clusters do not typically associate with other clusters (e.g., people in counties separated by hundreds of miles). In the majority of cluster-randomized trials, the clusters are physical places (e.g., schools, hospitals, counties, medical clinics). But the clusters can be groups of people such as families. Here we use the cluster terminology with the acknowledgements that the group term is equivalent and that clusters usually involve physical places. The latter consideration is practically important with regard to mobility because places are typically much less mobile than persons.

Another consideration in cluster-randomized prevention trials is whether a repeated cross-section design or a cohort (also known as longitudinal or panel) design is used. Repeated cross-section designs have no person mobility or attrition because the analysis considers only those persons who were present in the cluster (e.g., clinic, school) at a given time (cross-section), and thus include only one observed measure of any person (e.g., Localio et al. 2006; Stevens et al. 2005). Estimates of intervention effects on clusters (or groups) are based on whether aggregated negative outcome rates are lower in later cluster (or group) cross-sections than in the baseline cross-sections (Localio et al. 2006). Such analyses provide useful overall estimates of intervention effects when the focus is on the cluster or group as the unit of analysis.

It is also important to acknowledge that cluster randomization does not guarantee balance of potential person-level confounders between conditions (Giraudeau and Rivaud 2009; Localio et al. 2006; Pals et al. 2008), especially when there are a small or moderate number of clusters. In addition, estimates of intervention effects in repeated cross-section designs are susceptible to secular trends (Bauman et al. 1999). These could be "natural" reductions in the negative outcome that occur as population trends that have nothing to do with the preventive intervention. Such trends can lead to under-estimation of intervention effects (Bauman et al. 1999) without appropriate adjustment for such trends.

On the other hand, randomized trials that use *cohort* designs (sometimes known as longitudinal or panel designs), unlike repeated cross-section designs, collect data from the same persons over two or more time points. Tests for preventive intervention effects in cohort designs consider whether persons in control clusters develop negative outcomes more than those in intervention clusters. An important and generally recognized advantage in the analysis of cohort designs is that persons can "serve as their own controls" (e.g., Localio et al. 2006; Shadish et al. 2002). Regardless of what statistical method is applied, cohort designs provide a powerful design control for individual

differences and unmeasured covariates. This reduces the likelihood that any observed statistically significant difference between intervention and control groups was due to some unmeasured confounding variable rather than the intervention itself. Because cohort designs provide for this important statistical control, and thus stronger evidence for intervention effects, they are the subject matter for this analysis. Relatively few prevention trials have yet combined a repeated cross-section design with a smaller cohort design component (e.g., Brown et al. 2009) in order to take advantage of this cohort design feature. Separate reviews of cancer prevention and control trials from 2002 to 2006 (Murray et al. 2008), and two journals in public health and preventive medicine from 1998 to 2002 (Varnell et al. 2004), found that about 15% of such trials included both repeated cross-section and cohort components in their designs.

An additional advantage of cohort designs is the way they can take into account the effects of individual person characteristics (e.g., gender, ethnicity, family structure, IQ) to provide adjusted estimates of intervention benefit. In cohort designs the tests for covariate effects examine their impact on person-level change over time. That is important if a trial seeks to understand how an intervention impacts change in individual people. Such tests are not possible in repeated cross-section designs, which must focus on average changes in different clusters of people at each wave, especially baseline and endpoint or follow-up. In such designs, intervention effects adjusted for person characteristics can be done with post-hoc stratification of groups of persons with the same person-level characteristics. But the estimates of the effects of person-level characteristics, including baseline measures of the outcome, on the intervention will be less precise that that provided by a cohort design. Such precision may not be a priority in trials that emphasize average cluster-level change. In either case, inclusion of such person-level covariates can be important in estimating potential disparity effects—for example, is the trial more effective for males or females?

This study describes a classification based on the way traditional and emerging designs and analyses take into account the mobility of persons into and out of preventive trials. Our multilevel classification is structured according to two criteria. One is whether or not the design, analysis and interpretation are based on an *ITT* analysis or an analysis that takes into account *compliance* with the intervention protocol. The second criterion is whether the design and analysis of intervention effects focuses on the *persons* or the *clusters* within which persons are clustered. This involves careful consideration of whether person or cluster is the appropriate unit of inference (Boruch and Foley 2000; Donner and Klar 2004; Sherwin 1978). The classification reveals insights into current and future practices in the design and analysis of cluster-randomized cohort prevention trials. The classification is not intended to be completely exhaustive regarding all possible cluster-randomized designs. But it is important for any cohort trials that include cluster-randomization. More generally, this multilevel classification is a call for prevention scientists to think more systematically about the implications of person mobility for estimating intervention effects.

Our classification for addressing mobility is motivated by the need for a prevention scientist to make basic practical decisions in the design, analysis and interpretation of results from cluster-randomized cohort prevention trials:

1) Which persons should be included in the trial?
2) What data should be collected from persons and clusters and used in the analysis of the trial?
3) To what population can the trial results be generalized?
4) What is the appropriate analysis and interpretation of preventive intervention effects?

Elements of the classification proposed below are relevant to any cohort prevention trial, but our focus will be on cluster-randomized trials (Bloom et al. 1999; Boruch and Foley 2000; Donner and Klar 2004; Flay and Collins 2005; Murray 1998).

To provide some practical grounding for the classification, the next section presents a rationale for a closer examination of mobility as an important element in preventive interventions in one important context for children in the US – schools. It describes the research implications of students moving into and out of schools that are part of a preventive trial. That is followed by a presentation of our multilevel classification of designs for dealing with mobility with an emphasis on implications for internal and external validity.

Student Mobility as a Substantive and Methodological Issue

Person/student mobility is an inescapable fact of life for school-based and other cluster-based preventive trials. For example, in a nationally representative sample (Early Childhood Longitudinal Study-Kindergarten: US Department of Education, 2006), 35.9% of children changed schools between kindergarten and 3rd grade (National Research Council and Institute of Medicine 2010). In large city schools, student mobility can be incredibly high (e.g., Sampson et al. 2008). In Chicago city elementary schools, for example, mobility ranged from 0.3% to 66.7% in the 2001–02 school year. A 66.7% mobility rate means that two-thirds of the students who enrolled in one fall were not present the following fall. Even with a mobility rate of "only" 30%, 70% of the original students are still present after 1 year, 49% after 2 years, and only 34.3% after 3 years. Retaining at least 50% of the original students in a study over 4 years would require limiting research to schools with average mobility rates of less

than 16%. This would mean eliminating as many as 75% of Chicago schools (those at highest risk for behavioral and academic problems) from consideration. This would seriously affect the generalizeability of findings to populations with greatest need for effective interventions. This kind of mobility is problematic for interventions that are designed to have active participation, whether it involves taking a certain drug, exercising, or classroom activities, for several years.

Mobility rates raise important substantive considerations for school- and other cluster-based research. Their importance for school-based studies is demonstrated in Table 1, which shows how mobility rates correlate with other school-level variables in two large school districts. Mobility rates are highly correlated with ethnic distribution, poverty, attendance, class size, teacher quality, and student achievement on the SAT (Scholastic Aptitude Test). Indeed, it is a very strong correlate of achievement, along with poverty, ethnic distribution and attendance. Because mobility is associated with multiple risk factors, mobile students, whether entering or leaving a particular school, are also likely to be at higher risk of problem behaviors and low academic achievement than the average student (e.g., Eckenrode et al. 1995; Gruman et al. 2008; Pribesh and Downey 1999). As a result, in the presence of substantial mobility-related attrition, intervention effect estimates based only on those who remain in the schools apply to those at lower risk for problem behaviors.

Students Who Enter Study Schools – a Lost Group?

With heightened concerns about estimating intervention effects in the presence of attrition, most studies have overlooked students who entered study schools during a trial (late entrants). Little has been written (see Brown et al. 2008, 2009) about what, if anything, to do with these students when interventions are delivered to clusters over a period of time. Because of cluster-based intervention delivery, late entrants typically do receive part, if not most, of the intervention along with others in their cluster. Given their exposure to the intervention, concerns have been raised about whether they should be included in data collection and impact analyses or not (Brown et al. 2008, 2009). If they received most of the intervention, could they not in some way be included in estimates of intervention effects, even if they weren't in the school at baseline when the cluster-based intervention began?

It has been standard practice to exclude late entrants from data analysis in cluster-randomized controlled trials, although data have sometimes been collected from them to streamline data collection in group contexts (e.g., school classrooms). But such late entrant data is typically not used in the analysis of trial impact. The justifications for this are that they were not in a cluster (e.g., school) when it was randomly assigned to the intervention, so no baseline data are available for them, and those who joined the intervention condition after baseline (e.g., entered a school randomized to get the intervention) did not receive the same amount (dosage) of the intervention. These justifications are legitimate. However, it is likely that in some trials late entrants are not different on measured and unmeasured variables from those persons who were in the clusters originally randomized to receive the intervention. It is also possible that late entrants are not different from dropouts on measured and unmeasured characteristics. In either case, inclusion of late entrants could be justified because equality of intervention and control groups, originally induced by randomization, would be preserved. A similar reasoning is the basis for including late entrants as "refreshment" or "augmented" samples (e.g., Brown et al. 2008; Hirano et al. 2001) to preserve the characteristics of the original sample in longitudinal studies.

This logic leads to a practical reason for including late entrants in estimates of intervention effects – to retain specific characteristics of the original sample that the trial was designed to target and, thus, maintain external validity (Brown et al. 2008, 2009). This would reduce the need to use such procedures as "calibrating" the results of randomized trials to draw inferences about target populations (Frangakis 2009). There are other practical benefits for the inclusion of late entrants such as a need to maintain sample size so estimates of effects have adequate power (Hox 2010).

These considerations may outweigh the limitations associated with including late entrants in the analysis of cluster-randomized trials (Brown et al. 2008). Such limitations can be addressed when several waves of data are collected. Missing data techniques with informative auxiliary variables (e.g., Baker et al. 2006; Collins et al. 2001; Leon et al. 2007) can be used to allow absence of baseline and other waves of data (Brown et al. 2009). It is important to note that the use of informative auxiliary variables, those that accurately predict missingness, the response variable itself (Rubin 2006), and noncompliance (Jo and Muthen 2003; Jo and Stuart 2009; Roy et al. 2008) essentially provides a way of bringing mobility processes into the analysis of intervention effects. Differential dosage received, created in the intervention group by mobility, can be considered as a form of noncompliance. So the principle stratification analysis of noncompliance in randomized trials (Frangakis and Rubin 2002; Frangakis et al. 2002; Jo and Muthen 2003; Rubin 2006; Stuart et al. 2008) can be used to adjust estimates of intervention effects for mobility. Of course it must be acknowledged that there can be other forms of noncompliance as, for example, when teachers assigned to deliver an intervention choose not to do so.

Considering the possibility of including late entrants in trial analysis highlights the importance of collecting data on covariates that predict missingness, late entrants, dropouts, and stayers to make possible a statistical determination of whether there are significant differences between intervention

Table 1 School-level correlates with mobility and achievement[a]

Variable	Midwest schools		Southeastern schools	
	% mobility	SAT	% mobility	SAT
% mobility	1.00	−0.60	1.00	−0.72
% white	−0.37	0.70	−0.64	0.86
% free/reduced lunch	0.50	−0.79	0.72	−0.91
% parent involvement	−0.30	0.35	−0.41	0.51
% daily attendance	−0.53	0.69	−0.79	0.82
Average class size (grade 3)	−0.27	0.39	−0.06	0.17
% of classes taught by unqualified/inexperienced teachers	0.26	−0.39	0.41	−0.45
2002 Standardized Achievement Test (SAT) score	−0.60	1.00	−0.72	1.00

[a] Data are from two large city public elementary school districts. Midwestern data are from 2001 to 2002 for 483 schools Southeastern data are from 1996 to 1997 for 132 schools

and control clusters (Bloom et al. 1999; Boruch and Foley 2000; Brown et al. 2009; Flay and Collins 2005). If such differences exist, then interpretation of observed intervention effects must be adjusted accordingly. Protocols that assess covariates that predict attrition, noncompliance and the outcome itself can strengthen the claim that a cluster randomized trial had a statistically significant effect.

There may be some circumstances in which late entrants are different from stayers and dropouts on important confounding variables. For example, it is possible that families who move away from a low-income neighborhood because their financial fortunes are improving would be replaced by families whose financial fortunes are low or declining. Or, in neighborhoods undergoing ethnic compositional change, it is possible that dropouts are disproportionately from one ethnic group and late entrants from another. In such cases, dropouts and late entrants may be quite different, and such mobility could introduce differences between the population represented at baseline and the population represented at endpoint. In such cases it would be important to adjust estimated effects for ethnicity. Typically, such demographic changes are quite slow relative to the duration of most trials and, important from a methodological standpoint, are likely to be similar across well-matched randomly assigned treatment and control conditions.

In some trials the intervention itself can have an impact on either attrition, late entry or both. This issue is most well known in the context of differential attrition (e.g., Hewitt et al. 2010), which has long been recognized as a threat to the internal validity of a trial (e.g., Flay and Cook 1981; Shadish et al. 2002). Some interventions have negative characteristics, such as undesired side effects or no apparent benefit, that results in more persons from the intervention group leaving the trial than those from the control group.

When this occurs, it limits the kinds of conclusions that can be made from the trial. But if intervention and control groups after attrition are still equivalent on pre-randomization confounder covariates, this provides some evidence that the attrition was not selective. The mirror image of the differential attrition problem is differential late entry. An example of this would be if it was publicized that an intervention in specific clusters (e.g., schools, hospitals) was successful (in raising math scores, reducing delinquency, or curing cancer) resulted in persons moving into such successful clusters. If those late entrants had different confounder characteristics than who were in the baseline intervention group, then the intervention and control groups may no longer be equivalent on confounder characteristics. As with differential attrition, the extent to which this is a problem depends on whether the two conditions remain balanced on potential confounders.

The next section presents our multilevel classification as four approaches to answering the above questions. It provides methodological guidance for dealing with mobility, and offers recommendations for the design and analysis of future prevention trials (see Table 2 for a summary). As indicated above, the structure of the classification is based on two elements: 1) whether the analysis uses an ITT analysis strategy or a strategy that takes into account compliance with the intervention, and 2) whether the focus of the trial is on the *person* or the *cluster*. To provide some grounding for these approaches, they are described in the context of a school-based cluster-randomized cohort trial (waves of measurement nested within children nested within schools). Estimates for intervention effects with such data are easily accomplished with multilevel models (e.g., Donner and Klar 2000; Foshee et al. 2005; Murray 1998; Rabe-Hesketh and Skrondal 2008; Raudenbush and Bryk 2002). To streamline the presentation, we do not address the natural mobility of students from primary to middle to high schools that can be accommodated in our approaches (e.g., Goldstein et al. 2007).

The Person-Focused, Intent-To-Treat Approach

One approach to estimating intervention effects is ITT analysis that focuses on all students who were in study

Table 2 The multilevel classification of approaches to mobility

Focus of the Trial	Nature of the Analysis	
	Intent-To-Treat	Compliance
Person	1. Person-focused intent-to-treat approach: • Uses data from all persons originally in the clusters assigned to conditions. • Focus is on estimating the person level of program effect. • Persons who leave clusters (e.g., schools) during the trial are followed • Persons who enter research clusters during the trial are not assessed (or analyzed). • *Question answered*: What is the impact of the intervention assignment on persons when it is implemented under real-world conditions (including mobility)?	3. Person-focused compliance approach: • Focus is estimating the effect on persons who comply with the intervention. CACE analysis is preferred to "as-treated " analysis. • Dropouts are not followed. • Late entrants are not assessed (or analyzed). • *Question answered*: What is the impact of the intervention on those persons who complied with it to receive an adequate "dosage?"
Cluster	2. Cluster-focused intent-to-treat approach: • Uses data from all clusters originally assigned to conditions. • Focus is on estimating the cluster-level intervention effect – i.e., prevalence not incidence. • Requires assessing persons who enter research clusters during the trial. • Persons who leave research clusters during the trial are not followed. • *Question answered*: What is the impact of intervention assignment on clusters of persons?	4. Cluster-focused compliance approach: • Focus is only on those clusters that stay in the trial or in which the intervention is fully implemented. • *Question answered*: What is the impact of the intervention on clusters that complied with it?

schools at the beginning of the trial – when random cluster assignment was done. This requires following and collecting data from students at the endpoint who leave study schools during the trial. Students who enter study schools during the trial are not added to the study. As described above, following students who leave schools to the endpoint of the trial may be a challenge that results in substantial missing data at endpoint. The resulting non-missing endpoint sample that can be included in this analysis may be biased systematically in the direction of not including the highest-risk students who tend to leave study schools. However this problem can be addressed by an intensive effort to collect data that actually predict attrition, and using it in imputation models or related techniques that approximate intervention effects had there been no dropouts (Brown et al. 2009).

The research question that is answered with this approach is of the form, "What is the impact of assignment to receive the intervention on students when it is implemented in real-world conditions, which includes mobility?" We will refer to this as the "*person-focused intent-to-treat*" approach. The "intent-to-treat" idea as applied here means that the focus is on intervention impact on all students who were in study schools at the time of assignment (of schools) to condition. The corresponding policy question is of the form, "How does assignment to receive the intervention affect the health of the targeted students, whether or not they stay in the same school or move away, and whether or not they comply with the intervention?" With schools randomized to conditions, such analyses can produce valid and unbiased estimates of intervention effects on individual students if intervention and control schools, and corresponding sets of followed students, remain equivalent on potentially confounding variables (Borman and Dowling 2006; Brown et al. 2009). However, the interpretation of results is complicated by the fact that there may be a range of levels of dosage of the intervention for stayers and dropouts, neither of whom receive the full intervention. This form and other forms of noncompliance are not considered in an ITT analysis. In addition, for interventions that seek changes at the school level (Beets et al. 2008; Stevens et al. 2005) as well as the student level, the intervention exposure lost by students who leave intervention schools includes loss of exposure to school-wide components as well as classroom components.

The person-focused ITT approach requires special attention to attrition when whole schools are the unit of random assignment to conditions, especially when the intervention seeks changes in the schools, for example, involving changes to the learning climate at the school (e.g., Beets et

al. 2008; Bonell et al. 2010). In this case the unit of inference (Donner and Klar 2004) should be the school, and a cluster-focused approach (see below) is more appropriate. Random assignment of schools does not *guarantee* that students in the control and intervention conditions will have the same pre-randomization measured and unmeasured characteristics that might influence outcome scores (Brown et al. 2009; Giraudeau and Rivaud 2009; Localio et al. 2006; Pals et al. 2008), especially when the number of schools in each condition is small. Thus, there could be differences in control and intervention conditions that compromise interpretation of endpoint condition differences.

The person-focused, ITT approach to cluster-randomized trials typically ignores cluster characteristics, except for condition assignment, as potential causes of outcomes. It often includes person characteristics as pre-treatment covariates that can be used to provide estimates of intervention effects adjusted for person characteristics, and tests for whether the effect of the intervention differs by person characteristic (e.g., gender) as covariate-by-condition-by-time interaction terms (e.g., Brown et al. 2009). The focus on person characteristics in this approach can be extended to random coefficient models that allow change trajectories to vary across persons (e.g., Brown et al. 2009; Murray 1998; Murray et al. 2006). This focus on baseline person characteristics, and allowing intervention effects to vary across persons, are two features of a statistical model that can apply the person-focused ITT approach to mobility in cluster randomized trials.

The Cluster-Focused, Intent-To-Treat Approach

The enduring concern over attrition in randomized trials is based on the assumption that the key research question concerns program effects on individual students; that is, it is focused on individuals and is thus "person-focused." This is most appropriate when individuals are randomly assigned to conditions. But what is most appropriate when randomization is at the level of the school, and when the school is the entity with which a program intervenes (Boruch and Foley 2000; Cook 2005; Donner and Klar 2004; Flay and Collins 2005; Sherwin 1978)?

In some disciplines this issue is addressed largely through the distinction between repeated cross-section and cohort designs (e.g., Brown et al. 2009). The unit of inference in a repeated cross-section design is the cluster. In the cohort design it is the person, as all persons are followed to endpoint, and a person-focused ITT analysis is done (Brown et al. 2009). But there are other possibilities. Here we consider a cohort design where the unit of inference is the cluster. The practical implications are that dropouts are not followed and late entrants are included.

When whole schools are assigned to conditions, the key research question can be framed at the school level as, "What is the impact of school assignment to receive the intervention on the behavior of persons who attend the school?" We refer to this as the "*cluster-focused, intent-to-treat*" approach. In a school-based disease prevention study, this approach would focus on the prevalence or incidence of the disease in the school, not the incidence rate of the disease for individual students who were at one time in the school. In this approach, the cluster becomes the *unit of inference* (Donner and Klar 2004). This is especially appropriate because cluster is also the unit of randomization (Giraudeau and Rivaud 2009). Although data are collected on persons within clusters and analyzed as individual units (i.e., not aggregated), the cluster is the focus of analysis and generalization (Cook 2005; Donner and Klar 2004). This distinction is similar to the distinction between population-average and person-specific estimates (e.g., Crouchley and Davies 1999) of intervention effects.

The inclusion of cluster as an element in our classification, exemplified in the cluster-focused ITT and cluster-focused compliance approaches, promotes increased attention to cluster characteristics in estimating intervention effects in cluster-randomized trials. The cluster level analysis should collect cluster-level variables that predict cluster attrition so that imputation methods can be used to allow the use of data from clusters that drop out of the trial. Inclusion of cluster-level covariates can also be used in person-focused ITT analyses (e.g., Brown et al. 2009). However, a concern for trials with relatively few clusters is that each cluster-level covariate included in the model reduces the degrees of freedom available for tests of intervention effects in cluster-randomized trials by 1.

In the cluster-focused ITT approach, the appropriate data consist of repeated measures of persons in the same clusters (e.g., schools, counties) over time, where most of the same persons are in the same clusters for the duration of the trial. Every person in each cluster at each time of assessment is measured. Data from every assessed person are included in the analysis. This approach includes data from persons who leave intervention or control clusters ("dropouts") collected from them while they were in the intervention clusters, data from persons who enter intervention or control clusters ("late entrants") collected from them after they enter intervention or control clusters, data from persons who stay in intervention or control clusters for the duration of the study ("stayers"), as well as persons in intervention or control clusters who dropped out of the intervention ("cluster dropouts"). Valid estimates of intervention effects are obtained and are generalizable to clusters with similar levels of mobility. This analysis does not involve aggregation (Donner and Klar 2004; Taljaard et al. 2008).

To include late entrants, it is important to avoid informative assignment that could introduce imbalance on confounders in the intervention and control conditions (Brown et al. 2008). That means that the late entrants should have the same characteristics as those dropped out, and there was not differential attrition. In school-based trials, residence usually determines which school a child attends, and parent (or parent surrogate) mobility into and out of residences is based on a wide variety of considerations independent of a school participating in an intervention. That is a basis for trials including late entrants (e.g., Aber et al. 2003; Brown et al. 2008; Flay et al. 2004; Jones et al. 2011).

The interpretation of impact estimates from cluster-focused ITT analyses ignores any noncompliance at person and cluster levels. Just as with the person-focused approach, interpretation is easier if mobility rates are similar across conditions and the same types of persons move in and out of intervention and control clusters. In general, stronger intervention impact is expected in clusters with lower mobility, and weaker impact expected in clusters with higher mobility, as low mobility would be associated with more exposure to and compliance with the intervention.

The Person-Focused, Compliance Approach

A central feature of ITT analysis is that it estimates the effect of *assignment* to condition and ignores person or cluster attrition and compliance with the intervention or control condition. This is a practical approach that is easily justified. It is reasonable to test whether the offering of an intervention gets results, regardless of what attrition, compliance and beneficial effect processes may occur. Indeed, the basic ITT analysis provides unbiased estimates of intervention effects as long as there is no attrition from the trial, or attrition is taken into account with imputation or related techniques. But in the interest of developing better interventions, it is also important to determine whether the intervention was effective for those who actually complied with it – the complier average causal effect or CACE (Frangakis and Rubin 2002; Jo and Muthen 2003; Jo et al. 2008; Little et al. 2009; Stuart et al. 2008). The person-level CACE analysis requires a measure of compliance for each person within a cluster assigned to the intervention. This can be assessed in a variety of ways, depending on the nature of the trial. If the trial requires those assigned to the intervention to take a medicine, a blood test could determine whether the person was taking the medicine. If the trial involves persons attending group sessions, then attendance (number of sessions) could be used as a basis for a measure of compliance (e.g., Jo and Muthen 2003; Stuart et al. 2008). Compliance at the cluster (e.g., school, clinic) level could be measured in terms of how often trial components were delivered, or the quality of the delivery of trial components.

The logic of the CACE analysis is that a better test of a causal effect requires consideration of what outcome scores would have been had those persons assigned to the control condition been offered and complied with the intervention (e.g., Frangakis and Rubin 2002). If a series of assumptions about person response to being offered the treatment are plausible in a given trial, then an estimate of the intervention effect on those who complied with the intervention is possible. The details of the CACE analysis in randomized trials are well-described in this journal (Jo et al. 2010; Stuart et al. 2008) and extensively elsewhere (e.g., Frangakis and Rubin 2002; Jo and Muthen 2003; Jo et al. 2008; Little et al. 2009; Taylor and Zhou 2009). In our multilevel classification we refer to person-focused and place focused *compliance approaches* as any design and analysis combination that explicitly collects and analyzes appropriate data to take into account whether or not a person, or a cluster, complies when a cluster is assigned to an intervention condition.

To provide context for the person-focused compliance approach, it is useful to consider problems inherent in the analysis of data from cluster-based studies that analyze *only* those persons who stay in the intervention clusters for the life of the trial (stayers) – often called the "as-treated" or "on-treatment" analysis. Dropouts are not followed and late entrants are not assessed. The research question that can be answered with this analysis is of the form, "What is the impact of assignment to the intervention on those persons who stayed in intervention clusters to receive the full intervention?" With negligible attrition and high compliance, a basic pharmaceutical clinical trial analysis suffices (Brown et al. 2008). But as trials diverge from these circumstances, effect estimates can become suspect. Randomization or compliance may break down, leading to the terms "unhappy randomization" and "broken randomized experiment" (Barnard et al. 2003; Shadish et al. 2002). The results from an as-treated analysis can be generalized only to stable populations with low mobility. In the US, many causes of mobility such as homelessness, unemployment, and family disruption are associated with increased risk of negative outcomes. Thus, trial dropouts and late entrants are often at elevated risk for negative outcomes and in greatest need of the benefits of the intervention. To address such disparity in the estimation of intervention effects, it is important that analyses retain the more mobile segments of the population. Designs that allow late entry and analyses that include those persons who do not comply can improve the inclusion of mobile segments of a population.

The analysis of data from a person-focused compliance design allows generalizations that include appropriate consideration of persons who do not comply with trial conditions. Attrition and late entry can be addressed, as described above,

using multiple imputation or maximum likelihood estimation (Brown et al. 2008; Collins et al. 2001; Foshee et al. 2005; Kellam et al. 2008), especially when covariates that predict missingness, and the outcome itself, are collected and used in the analysis (Baker et al. 2006; Collins et al. 2001; Jo and Muthen 2003; Leon et al. 2007; Roy et al. 2008; Rubin 2006). An effective CACE analysis for the person-focused compliance approach requires a design protocol that collects a person-level measure of compliance, and person-level covariates that predict compliance (Jo 2002; Jo and Muthen 2003; Roy et al. 2008; Rubin 2006). Indeed, the availability of such person-level covariates reduces the reliance of a CACE analysis on untestable assumptions (Jo 2002; Jo et al. 2010).

The Cluster-Focused Compliance Approach

Clusters as well as persons can drop out of a trial, enter it late, or fail to comply with the intervention. When that happens in cluster-randomized trials, it can threaten the internal validity as well as generalizability of the estimates of intervention effects, as described above. In some cluster-based trials, the analysis simply excluded any clusters that did not implement the intervention well (e.g., Battistich et al. 2000), in an as-treated analysis. Intervention effect estimates then answer the question, "What is the impact of the intervention in clusters where it is implemented well?" This can be useful information. But questions remain about whether a break in the condition balance established by cluster randomization introduced confounder effects; i.e., important differences in characteristics of control and intervention clusters. Such differences can become masked as intervention effects in the analysis, but could be reduced in a matched-pair design by dropping matched pairs of clusters rather than just individual clusters (Battistich et al. 2000). In addition, dropping non-complying clusters from the intervention group may limit generalizations to the population targeted by the trial. Clusters that did not comply may be those that most needed the intervention, thus creating a disparity bias in effect estimates—e.g., the trial proved to be effective, but only for low-risk clusters. Because of these limitations, an as-treated analysis should be secondary to analysis using the cluster-focused ITT approach described above.

The *cluster-focused compliance approach* in our classification seeks to avoid these threats to internal and external validity by taking attrition and compliance into account at the cluster level using appropriate covariates. Similar to the person-focused compliance approach described above, this can be done with missing data techniques such as maximum likelihood estimation and multiple imputation, and for non-compliance analysis, the potential outcomes models. However, the important difference in the cluster-focused compliance approach is that models for missingness and noncompliance focus on understanding why *clusters* drop out of trials (e.g., stop providing data), why they may enter late, and why they do not comply with the intervention protocol (e.g., data were provided but there was no compliance). Such understanding can provide the basis for collecting covariates that are *characteristics of the clusters* that can be used in the analysis of missingness and noncompliance. Few studies have collected detailed data on cluster-level compliance (e.g., Aber et al. 2003; Beets et al. 2008). In cluster-randomized cohort trials, clusters are much less likely than persons to drop out of a trial or enter late. So missingness analysis is usually less of a problem in the cluster-focused than the person-focused approach. However, compliance with an intervention may vary widely across clusters (e.g., Aber et al. 2003; Beets et al. 2008). In such cases good measures of cluster-level compliance, and covariates that predict cluster-level compliance, provide the basis for a CACE analysis of the effectiveness of a trial on clusters that comply with it.

The cluster-focused compliance approach represents a conceptual shift compared with the person-focused compliance approach, as it does in the cluster-focused ITT approach described above. The cluster becomes the *unit of inference* (Donner and Klar 2004), even though measures of individual persons, rather than cluster aggregates, are included in the multilevel analysis, and the focus of interest in the trial is on groups of persons rather than on individual persons.

Application Issues

A key point in the multilevel classification is that when person mobility creates substantial attrition, late entry or noncompliance, accurate estimates of intervention effects require the collection of covariates associated with these phenomena. Although the statistical methods for utilizing such covariates in estimates of trial effects have already been developed, research has as yet only begun to identify appropriate covariates (e.g., Baker et al. 2006; Jo and Muthen 2003; Leon et al. 2007; Roy et al. 2008; Rubin 2006). This is a challenge for prevention scientists who study populations with person mobility. It adds a burden to a protocol to not only collect data on potential confounder variables to insure balance between conditions, but to also collect covariates associated with missingness, noncompliance, and the outcome itself. As knowledge of these covariates expands, estimates of trial effects where person mobility is present will improve because mobility processes will essentially be included in the model.

The multilevel mobility classification is not primarily statistical in nature, but some statistical techniques are particularly well-suited for these approaches. The 3-level

hierarchical linear mixed model, and the latent growth curve mixture model (for noncompliance) provide convenient formats for representing each approach to mobility (e.g., Foshee et al. 2005; Jo and Muthen 2003; Jo et al. 2008; Jones et al. 2011; Muthen et al. 2002; Raudenbush and Bryk 2002). The generalized linear mixed model (e.g., Rabe-Hesketh and Skrondal 2008; Raudenbush and Bryk 2002) provides the same format for different outcome types, such as binary, counts, durations, and ordinal. With the nested data structure (waves nested in persons nested in clusters) the multilevel longitudinal model tests for a difference in the trajectory of change in the negative health outcome between intervention and control groups from baseline to endpoint. Evidence of a beneficial preventive intervention is found if the growth in the negative outcome is significantly lower in the intervention clusters than in the control clusters (a time-by-condition interaction). Attrition and late entry are addressed either by imputing missing values before estimating the models (e.g., Foshee et al. 2005; Taljaard et al. 2008), or including predictors of missingness in the models using an inclusive technique (Collins et al. 2001). The multilevel formulation also applies for the CACE noncompliance analysis. This can be done in a growth mixture model where the sample is considered as a mixture of compliers and noncompliers (e.g., Jo and Muthen 2003; Jo et al. 2008).

Conclusions

This commentary presents a systematic multilevel classification for conceptualizing and analyzing cluster-randomized cohort prevention trials in which person mobility creates attrition, late entry, and noncompliance. Novel features of this classification are that: 1) its multilevel structure highlights the importance of cluster as a unit of inference (Cook 2005; Donner and Klar 2004) and how cluster-level characteristics can be integrated into designs and analyses that contribute to understanding variation in the effects of an intervention, 2) it extends previous work on late entrants (Brown et al. 2008, 2009) and reveals new advantages associated with including late entrants in cluster-randomized trials, 3) it demonstrates how a focus on cluster as a unit of inference (Donner and Klar 2004) leads to a new type of ITT analysis – the cluster-focused ITT approach, and 5) it highlights the importance for accurate imputation and noncompliance analysis of collecting information on informative person- and cluster-level covariates that are associated with the outcome measure, missingness, and compliance with the intervention.

Recent work by some prevention scientists has suggested that accurate estimates of intervention effects in some studies requires extensions to the traditional ITT design and analyses for many preventive trials (e.g., Brown et al. 2008, 2009). For example Brown et al. (2008) made a distinction between "pharmaceutical randomized clinical trials," which established the ITT criteria years ago, and "randomized field trials" that are increasingly being used in prevention science. The widespread application of cluster-randomized trials and multilevel analysis has created a variety of options for such field trials, as well as challenges for design, analysis and interpretation. This commentary extends an emphasis on conceptual, protocol and validity issues, in a Campbellian-style analysis (Campbell 1957; West et al. 2008), to the problem of person mobility in cluster-randomized trials. Our multilevel classification provides a systematic organization of primary options for taking mobility into account in cluster-randomized cohort preventive trials. It seeks to draw renewed and more focused attention to design and analysis issues that can improve the accuracy of inferences about intervention effects.

In our view, too often the emphasis in drawing conclusions about randomized prevention trials in the presence of person mobility has been simply on "Was the intervention effective?" without considering "For what populations was the trial effective?" Too often, conclusions have been framed as, "the intervention prevented a negative outcome" rather than "the intervention prevented a negative outcome for individuals" or "the intervention reduced the rate of a negative outcome for clusters." Too often, valuable data on those who dropped out of a trial, and on those who were late entrants, have been discarded.

Any of the four approaches to mobility described here can be used to draw valid conclusions about the effectiveness of preventive interventions under certain conditions. The person-focused ITT approach, of course, is the most well-known. Its reliance on assessing all persons at endpoint who were randomized at baseline can be problematic in at-risk populations because they are often more mobile and difficult to find at endpoint. Imputation and related techniques can offer solutions if attrition is not high and informative covariates are available. An advantage of the cluster-focused ITT approach is that clusters (e.g., schools, counties, clinical, hospitals) are typically less mobile than persons, so that both endpoint and baseline assessments are more likely to be available. This approach shares the advantage of the person-focused ITT approach to randomized cohort trials that persons serve as their own controls, which provides a stronger basis for causal inference. It should be noted that imputation methods and software are available to do multilevel imputation that considers both person and cluster levels (e.g., Carpenter et al. 2011; Yucel 2008).

Noteworthy in practical terms, and consistent with Brown et al. (2008), our mobility analysis found that the cluster-focused ITT approach has largely unrecognized

advantages for estimating ITT effects of preventive interventions in the presence of mobility. A key issue, in that respect, is the inclusion of late entrants in an ITT analysis when multiple waves of response data are collected in a cluster-randomized trial. Inclusion of late entrants limits the impact mobility processes can have on distorting conclusions about intervention effects by retaining balance between conditions, thus contributing to internal validity. In addition it serves to retain the characteristics of a sample that reflects the population being targeted by the trial, thus improving external validity.

The cluster-focused ITT approach meets ITT criteria by focusing on the cluster rather than the person. However, in using the cluster-focused ITT approach, researchers must give careful attention to the person and cluster characteristics of their sample over time in order to detect any changes from the baseline. When reporting results they need to specify clearly the unit of inference (Donner and Klar 2004). With the cluster-focused ITT approach, that unit is the cluster (e.g., school, clinic, family) rather than the person. The person-level data are used in order to increase the basis for causal inference, and provide a basis for estimates adjusted for potential disparity effects as well as potential moderator or mediator analysis (Fairchild and MacKinnon 2009).

To date, only a few researchers have published results from preventive trials using the cluster-focused ITT approach. Flay and colleagues (Flay et al. 2004) reported results from a randomized trial of the Aban Aya Youth Program in Chicago and Jones and colleagues (2011) reported results from a school-based cluster-randomized trial in New York City using the cluster-based ITT approach. Both of these studies were conducted in inner-city school districts with high levels of student mobility. This approach produces valid estimates of intervention effects as long as mobility is not affected by the intervention. Brown et al. (2008) and the Multisite Violence Prevention Project (2008) describe variations of this approach.

There are limitations to the multilevel mobility classification system presented here. It considers only cluster-randomized trials and not those based on randomization at the person level. Also, it uses the basic distinction between ITT and compliance models without addressing circumstances where hybrid models may be appropriate (e.g., Brown et al. 2009; Jo et al. 2008; Raudenbush and Bryk 2002). In addition, it did not examine power issues in these models. The number of clusters used in these multilevel designs and models needs to be large enough to provide adequate statistical power to detect intervention effects (Donner and Klar 2004; Hox 2010). That number depends on characteristics such as the intraclass correlation, the effect size, and the number of random effects in the model. However, Bayesian estimation, using such methods as Markov Chain Monte Carlo methods, can require far fewer clusters (e.g., Frangakis et al. 2002; Hox 2010; Localio et al. 2006).

Progress toward more accurate estimates of preventive intervention effects will require closer integration of design and statistical methods, as promoted in this paper. Also important in this regard are efforts toward integration of the diverse statistical techniques currently used to address the problems associated with mobility (e.g., Baker et al. 2006; Brown et al. 2008; Taylor and Zhou 2009). In the context of mobility in cluster-randomized trials, the statistical options have advanced more quickly than prevention science's ability to fully utilize them. That will require a much better understanding of why attrition, late entry and noncompliance occur in randomized field trials.

Acknowledgments Preparation of this paper was supported by grants to each of the authors as part of the Social and Character Development (SACD) Research program funded by the Institute of Education Sciences (IES), U.S. Department of Education. The content of this publication does not necessarily reflect the views or policies of IES or the U.S. Government.

References

Aber, J. L., Brown, J. L., & Jones, S. M. (2003). Developmental trajectories toward violence in middle childhood: Course, demographic differences, and response to school-based intervention. *Developmental Psychology, 39*, 324–348.

Baker, S. G., Fitzmaurice, G. M., Freedman, L. S., & Kramer, B. S. (2006). Simple adjustments for randomized trials with nonrandomly missing or censored outcomes arising from informative covariates. *Biostatistics, 7*, 29–40.

Barnard, J., Frangakis, C. E., Hill, & Rubin, D. T. (2003). Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. *Journal of the American Statistical Association*, *98*, 299–311.

Battistich, V., Schaps, E., Watson, M., Solomon, D., & Lewis, C. (2000). Effects of the Child Development Project on students' drug use and other problem behaviors. *The Journal of Primary Prevention, 21*, 75–99.

Bauman, K. E., Suchindran, C. M., & Murray, D. M. (1999). The paucity of effects in community trials: Is secular trend the culprit? *Preventive Medicine, 28*, 426–429.

Beets, M. W., Flay, B. R., Vuchinich, S., Acock, A. C., Li, K., & Allred, C. (2008). School climate and teachers' beliefs and attitudes associated with implementation of the Positive Action program: A diffusion of innovations model. *Prevention Science, 9*, 264–275.

Bloom, H. S. (2005) Randomizing groups to evaluate place-based programs. In H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (pp. 115–172). New York: Russell Sage Foundation.

Bloom, H. S., Bos, J. M., & Lee, S.-W. (1999). Using cluster random assignment to measure program impacts: Statistical implications for the evaluation of education programs. *Evaluation Review, 23*, 445–469.

Bonell, C., Sorhaindo, A., Strange, V., Wiggins, M., Allen, E., Fletcher, A., et al. (2010). A pilot whole-school intervention to improve school ethos and reduce substance use. *Health Education, 110*, 252–272.

Borman, G. D., & Dowling, N. M. (2006). Longitudinal achievement effects of multiyear summer school: Evidence from the Teach

Baltimore randomized field trial. *Educational Evaluation and Policy Analysis, 28*, 25–48.

Boruch, R. G., & Foley, E. (2000). The honestly experimental society: Sites and other entities as the units of allocation and analysis in randomized trials. In L. Bickman (Ed.), *Validity and social experimentation: Donald Campbell's legacy, Volume 1* (pp. 198–238). Thousand Oaks, CA: Sage Publications.

Brown, C. H., Wang, W., Kellam, S. G., Muthen, B. O., Petras, H., Toyinbo, P., et al. (2008). Models for testing and evaluating impact in randomized field trials: Intent-to-treat analyses for integrating the perspectives of person, place, and time. *Drug and Alcohol Dependence, 95S*, S74–S104.

Brown, E. C., Graham, J. W., Hawkins, J. D., Arthur, M. W., Baldwin, M. M., Oesterle, S., et al. (2009). Design and analysis of the Community Youth Development Study longitudinal cohort sample. *Evaluation Review, 33*, 311–334.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin, 54*, 297–312.

Carpenter, J. R., Goldstein, H. & Kenward, M. G. (2011). REALCOM-IMPUTE software for multilevel multiple imputation with mixed response types. *Journal of Statistical Software, 45*, 1–14.

Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *The New England Journal of Medicine, 357*, 370–379.

Clark, T. W., Pareek, M., Hoschler, K., Dillon, H., Nicholson, K. G., Groth, N., & Stephenson, I. (2009). Trial of 2009 Influenza A (H1N1) monovalent MF59-adjvanted vaccine. *The New England Journal of Medicine, 361*, 2424–2435.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in missing data procedures. *Psychological Methods, 6*, 330–351.

Cook, T. D. (2005). Emergent principles for the design, implementation, and analysis of cluster-based experiments in social science. *The Annals of the American Academy of Political and Social Science, 599*, 176–198.

Cornfield, J. (1978). Randomization by group: A formal analysis. *American Journal of Epidemiology, 108*, 100–102.

Crouchley, R., & Davies, R. B. (1999). A comparison of population average and random-effect models for the analysis of longitudinal count data with base-line information. *Journal of the Royal Statistical Society, Series A, 162*, 331–347.

Danese, A., Moffitt, T. E., Pariente, C., Poulton, R., Caspi, A. (2008). Elevated inflammation levels in depressed adults with a history of childhood maltreatment. *Archives of General Psychiatry, 65*, 409–415.

DeGarmo, D. S., Eddy, J. M., Reid, J. B., & Fetrow, R. A. (2009). Evaluating mediators of the impact of the Linking the Interests of Families and Teachers (LIFT) multimodal preventive intervention on substance use initiation and growth across adolescence. *Prevention Science, 10*, 208–220.

Dodge, K. A., Greenberg, M. T., Malone, S. M., & Conduct Problems Prevention Group. (2008). Testing an idealized dynamic model of the development of serious violence in adolescence. *Child Development, 79,* 1907–1927.

Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. London, England: Arnold.

Donner, A., & Klar, N. (2004). Pitfalls and controversies in cluster randomization trials. *American Journal of Public Health, 94*, 416–422.

Eckenrode, J., Rowe, E., Laird, M., & Brathwaite, J. (1995). Mobility as a mediator of the effects of child maltreatment on academic performance. *Child Development, 66*, 1130–1142.

Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention Science, 10*, 87–99.

Fitzmaurice, G. M., & Laird, N. M. (2000). Generalized linear mixture models for handling nonignorable dropouts in longitudinal studies. *Biostatistics, 1,* 141–156.

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine, 15*, 451–474.

Flay, B. R., & Collins, L. M. (2005). Historical review of school-based randomized trials for evaluating problem behavior prevention programs. *The Annals of the American Academy of Political and Social Science, 599*, 115–146.

Flay, B. R., & Cook, T. D. (1981). Evaluation of mass media prevention campaigns. In R. R. Rice & W. Paisley (Eds.), *Public communication campaigns* (pp. 239–313). Beverly Hills, CA: Sage.

Flay, B. R., Graumlich, S., Segawa, S., Burns, J. L., Holliday, M. Y., & Investigators, A. A. (2004). Effects of two prevention programs on high-risk behaviors among African-American youth: A randomized trial. *Archives of Pediatric & Adolescent Medicine, 158*, 377–384.

Foshee, V. A., Bauman, K. E., Ennett, S. T., Suchindran, C., Benefield, T., & Linder, G. F. (2005). Assessing the effects of the dating violence prevention program "Safe Dates" using random coefficient regression modeling. *Prevention Science, 6*, 245–258.

Frangakis, C. E. (2009). The calibration of treatment effects from clinical trials to target populations. *Clinical Trials, 6*, 136–140.

Frangakis, C. E., & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics, 58*, 1–29.

Frangakis, C. E., Rubin, D. B., & Zhou, X. H. (2002). Clustered encouragement design with individual noncompliance: Bayesian inference and application to advance directive forms. *Biostatistics, 3*, 147–164.

Giraudeau, B., & Rivaud, P. (2009). Preventing bias in cluster randomized trials. *PLoS Medicine, 6*, e1000065.

Goldstein, H., Burgess, S., & McConnell, B. (2007). Modelling the effect of pupil mobility on school differences in educational achievement. *Journal of the Royal Statistical Society: Series A, 170*, 941–954.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576.

Gruman, D. H., Harachi, D. W., Abbott, R. D., Catalano, R. F., & Fleming, C. B. (2008). Longitudinal effects of student mobility on three dimensions of elementary school engagement. *Child Development, 79*, 1833–1852.

Hewitt, C. E., Kumaravel, B., Dumville, J. C., & Torgerson, D. J. (2010). Assessing the impact of attrition in randomized controlled trials. *Journal of Clinical Epidemiology, 63*, 1264–1270.

Hirano, K., Imbens, G., Ridder, G., & Rubin, D. (2001). Combining panel data sets with attrition and refreshment samples. *Econometrica, 69*, 1645–1659.

Hox, J. (2010). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics, 27*, 385–409.

Jo, B., & Muthen, B. (2003). Longitudinal studies with intervention and noncompliance: Estimation of causal effects in growth mixture modeling. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues and applications* (pp. 112–139). Mahwah NJ: Lawrence Erlbaum.

Jo, B., & Stuart, E. A. (2009). On the use of propensity scores in principal causal effect estimation. *Statistics in Medicine, 28*, 2857–2875.

Jo, B., Asparouhov, T., Muthen, B., Ialongo, N., & Brown, C. H. (2008). Intention-to-treat analysis in cluster randomized trials with noncompliance. *Statistics in Medicine, 27*, 5565–5577.

Jo, B., Ginexi, E. M., & Ialongo, N. S. (2010). Handling missing data in randomized experiments with noncompliance. *Prevention Science, 11*, 384–396.

Jones, S., Brown, J., & Aber, J. L. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development, 82*, 533–554.

Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N., Wang, W., Toyinbo, P., et al. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence, 95*, S5–S28.

Leon A. C., Demirtas, H., & Hedeker, D. (2007). Bias reduction with an adjustment for participants' intent to drop out of a randomized controlled clinical trial. *Clinical Trials*, *4*, 540–547.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed). New York: Wiley.

Little, R. J., Long, Q., & Lin, X. (2009). A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics, 65*, 640–649.

Liu, L. C., Flay, B., & Investigators, A. A. (2009). Evaluating mediation in longitudinal multivariate data: Mediation effects for the Aban Aya Youth Project drug prevention program. *Prevention Science, 10*, 197–207.

Localio, A. R., Berlin, J. A., & Ten Have, T. R. (2006). Longitudinal and repeated cross-sectional cluster-randomization designs using mixed effects regression for binary outcomes: Bias and coverage of frequentist and Bayesian methods. *Statistics in Medicine, 25*, 2720–2736.

Multisite Violence Prevention Project. (2008). Impact of a universal school-based violence prevention program on social-cognitive outcomes. *Prevention Science, 9*, 231–244.

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.

Murray, D. M., Van Horn, M. L., Hawkins, J. D., & Arthur, M. W. (2006). Analysis strategies for a community trial to reduce adolescent ATOD use: A comparison of random coefficient and ANOVA/ANCOVA models. *Contemporary Clinical Trials, 27*, 188–206.

Murray, D. M., Pals, S. L., Blitstein, J. L., Alfano, C. M., & Lehman, J. (2008). Design and analysis of group-randomized trials in cancer: A review of current practices. *Journal of the National Cancer Institute, 100*, 483–491.

Muthen, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S.-T., Yang, C.-P., et al. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics, 3*, 459–475.

National Research Council and Institute of Medicine. (2010). *Student mobility: Exploring the impact of frequent moves on achievement: Summary of a workshop*. (Beatty, A., Rapporteur). Committee on the impact of mobility and change on the lives of young children, schools, and neighborhoods. Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Osann K., Wenzel L., Dogan A., Hsieh S., Chase D. M., Sappington S., Monk B. J., & Nelson E. L. (2011). Recruitment and retention results for a population-based cervical cancer biobehavioral clinical trial. *Gynecologic Oncology, 121,* 558–563.

Pals, S. L., Murray, D. M., Alfano, C. M., Shadish, W. R., Hannan, P. J., & Baker, W. L. (2008). Individually randomized group treatment trials: A critical appraisal of frequently used design and analytic procedures. *American Journal of Public Health, 98*, 1418–1424.

Pribesh, S., & Downey, D. B. (1999). Why are residential and school moves associated with poor school performance? *Demography, 36,* 521–534.

Prinz, R. J., Dumas, J. E., Smith, E. P., Laughlin, J., White, D., & Barrón, R. (2001). Recruitment and retention of participants in prevention trials. *American Journal of Preventive Medicine, 20 (Supplement)*, 31–37.

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata* (2nd ed.). College Station, TX: Stata Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Ribisl, K. M., Walton, M. A., Mowbray, C. T., Luke, D. A., Davidson, W. S., & Bootsmiller, B. J. (1996). Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: Review and recommendations. *Evaluation and Program Planning, 19*, 1–25.

Roy, J., Hogan, J. W., & Marcus, B. H. (2008). Principal stratification with predictors of compliance for randomized trials with two active treatments. *Biostatistics, 9*, 277–289.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rubin, D. B. (2006). Causal inference through potential outcomes and principal stratification. *Statistical Science, 21*, 299–309.

Sampson, R. J., Sharkey, P., & Raudenbush, S. W. (2008). Durable effects of concentrated disadvantage on verbal ability among African-American children. *PNAS, 105*, 845–852.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Wadsworth.

Sherwin, R. (1978). Controlled trials of the diet-heart hypothesis: Some comments on the experimental unit. *American Journal of Epidemiology, 108*, 92–99.

Stevens, J., Murray, D. M., Catellier, D. J., Hannan, P. J., Lytle, L. A., Elder, J. P., et al. (2005). Design of the Trial of Activity in Adolescent Girls (TAAG). *Contemporary Clinical Trials, 26*, 223–233.

Stuart, E. A., Perry, D. F., Le, H.-N., & Ialongo, N. (2008). Estimating intervention effects of prevention programs: Accounting for noncompliance. *Prevention Science*, *9*, 288–298.

Taljaard, M., Donner, A., & Klar, N. (2008). Imputation strategies for missing continuous outcomes in cluster randomized trials. *Biometrical Journal, 50*, 329–345.

Taylor, L., & Zhou, X. H. (2009). Multiple imputation methods for treatment noncompliance and nonresponse in randomized clinical trials. *Biometrics, 65*, 88–95.

U.S. Census Bureau. (2009). *Current population survey, annual social and economic supplement*. Table 1. General Mobility: 2008 to 2009. Washington, DC: US Census Bureau.

U.S. Department of Education, National Center for Education Statistics. (2006). *ECLS-K longitudinal kindergarten-fifth grade public-use data file and electronic codebook* (CD-ROM). (NCES 2006–035). Washington, DC: Author.

Varnell, S., Murray, D. M., Janega, J. B., & Blitstein, J. L. (2004). Design and analysis of group-randomized trials: A review of recent practices. *American Journal of Public Health, 94*, 393–399.

West, S. G., Duan, N., Pequegnat, W., Gaist, P., Des Jaríais, D. C., Holtgrave, D., et al. (2008). Alternatives to randomized controlled trials. *American Journal of Public Health, 98*, 1359–1366.

Yucel, R. M. (2008). Multiple imputation inference for multivariate multilevel continuous data with ignorable non-response. *Philosophical Transactions of the Royal Society A, 366*, 2389–2403.