Accounting for Cluster Randomization: A Review of Primary Prevention Trials, 1990 through 1993

ABSTRACT

Objectives. This methodological review aims to determine the extent to which design and analysis aspects of cluster randomization have been appropriately dealt with in reports of primary prevention trials.

Methods. All reports of primary prevention trials using cluster randomization that were published from 1990 to 1993 in the *American Journal of Public Health* and *Preventive Medicine* were identified. Each article was examined to determine whether cluster randomization was taken into account in the design and statistical analysis.

Results. Of the 21 articles, only 4 (19%) included sample size calculations or discussions of power that allowed for clustering, while 12 (57%) took clustering into account in the statistical analysis.

Conclusions. Design and analysis issues associated with cluster randomization are not recognized widely enough. Reports of cluster randomized trials should include sample size calculations and statistical analyses that take clustering into account, estimates of design effects to help others planning trials, and a table showing the baseline distribution of important characteristics by intervention group, including the number of clusters and average cluster size for each group. (*Am J Public Health.* 1995;85:1378–1383)

Judy M. Simpson, PhD, Neil Klar, PhD, and Allan Donner, PhD

Introduction

Primary prevention trials are designed to test interventions aimed at reducing the incidence of disease or ill health by altering susceptibility or reducing exposure for susceptible individuals. The intervention may involve health promotion (e.g., school-based drug use prevention programs or work-site nutrition intervention programs) or may consist of a specific protective measure (e.g., immunization or fluoridation of the water supply).[1] Such interventions can be aimed either at the individual or at an entire community.

Whereas in treatment trials it is common practice to randomize individuals to treatment groups, it is not always feasible to do so in primary prevention trials. The most obvious example is the community intervention trial, in which the intervention itself is delivered to a whole community and aspects of the community infrastructure, such as mass media, are used. In other studies, the intervention is delivered to individuals but may take advantage of some social grouping, such as school classes, work sites, or, particularly in developing countries, villages. In either case, it is not the individual who is randomly allocated to an intervention group, but a cluster of individuals.

An important feature of cluster randomized trials is that individuals within a cluster, and therefore receiving the same intervention, will probably resemble each other more than individuals in different clusters, so their responses to the intervention will not be statistically independent. This could arise because of inherent similarities among cluster members or because of some feature unique to the cluster, such as the class teacher or workplace management. Another possibility is that social interaction between members of the same cluster, such as classmates or workmates, leads them to respond similarly. If the outcome is an infectious disease, within-cluster similarities may occur as a result of the spread of the disease among cluster members. In all cases, the between-cluster variation in response will consequently be greater than that within clusters.

If subjects are randomized in clusters rather than individually, this aspect of the trial must be taken into account in both the design and the statistical analysis. Within-cluster similarities in response decrease the effective sample size in comparison with that required when using individual randomization. Furthermore, the statistical methods used in the analysis must take into account the correlation among responses of individuals in the same cluster.

In this article, we report the results of a methodological review of primary prevention trials in which cluster randomization was used. The trials reviewed were published between 1990 and 1993 in two major journals in which the results of such trials are commonly reported. Our aim was to determine the extent to which design and analysis aspects of cluster randomization were appropriately dealt with in reports of these trials.

Methods

Manual searching was used to identify all reports of primary prevention trials

Judy M. Simpson is with the Department of Public Health, University of Sydney, Sydney, New South Wales, Australia. Neil Klar and Allan Donner are with the Department of Epidemiology and Biostatistics, University of Western Ontario, London, Ontario, Canada.

Requests for reprints should be sent to Judy M. Simpson, PhD, Department of Public Health, A27, University of Sydney, 2006, New South Wales, Australia.

This paper was accepted February 24, 1995.

using cluster randomization that were published from 1990 to 1993 in two major journals: the *American Journal of Public Health* and *Preventive Medicine.* An article was considered eligible if it reported the results of a trial of a primary prevention measure, clusters of individuals were randomly allocated and not merely assigned to a particular intervention group, and the unit of observation was the individual, while the unit of randomization was the cluster. If it was not clear whether an article was eligible, it was omitted; for example, one article was excluded because it did not contain a clear statement about randomization. In some studies, individuals were randomized and then received the allocated intervention as a group. These studies were not included, although the group dynamics may have induced correlations between individual responses, which would raise analytic issues similar to those raised by the actual randomization of clusters.

The search identified 22 eligible articles, 13 in the *American Journal of Public Health*[2–14] and 9 in *Preventive Medicine.*[15–23] Two of these articles, one in each journal, gave results of the same trial at different follow-up times.[11,23] Because the methods used in the two articles were the same, we decided to include only the one published in *Preventive Medicine* in order to have more equal numbers from the two journals.

Each article was examined independently by all three authors to determine whether it satisfied two selected methodological criteria that are especially relevant to the design and analysis of cluster randomized trials. Disagreements were resolved by roundtable discussion among all three authors.

Between-Cluster Variation Accounted for in Sample Size and/or Power Calculations

In a cluster randomized trial with c clusters and m individuals per cluster, the variance of the mean intervention effect for each group is as follows:

$$\frac{\sigma^2/m + \sigma_c^2}{c}, \qquad (1)$$

where σ^2 is the within-cluster variance of the response (i.e., the mean squared deviation of individual responses within a cluster from the mean response for that cluster) and σ_c^2 is the between-cluster variance (i.e., the mean squared deviation of the mean responses in the clusters from the overall mean).[24] If the between-cluster variation is ignored, the variability of the intervention effect is underestimated as

$$\frac{\sigma^2 + \sigma_c^2}{mc}. \qquad (2)$$

Sample size calculations that ignore clustering thus underestimate the required numbers.[25]

One can compensate for clustering by multiplying the sample size by a "design effect" factor, $1 + (m - 1)\,\rho$, where ρ is the intracluster correlation coefficient, given by

$$\rho = \frac{\sigma_c^2}{\sigma^2 + \sigma_c^2}, \qquad (3)$$

that is, the ratio of between-cluster variability to total variability. If all members of a cluster respond identically ($\sigma^2 = 0$), then $\rho = 1$; if there is no variability between clusters ($\sigma_c^2 = 0$), then $\rho = 0$ and the design effect is 1. The design effect is also 1 when $m = 1$, because then individuals are being randomized. The design effect is also known as the variance inflation factor because it represents the amount by which the variance estimate obtained if clustering is ignored (given in expression 2) needs to be inflated in order to obtain the correct variance in expression 1. From expression 1, it can be seen that it is more effective to increase the number of clusters, c, rather than the number of individuals per cluster, m, but this may be difficult if the number of available clusters is limited and expensive if the cluster size is large. Note that the effective sample size of a study (i.e., the equivalent number of independent observations) is the actual size divided by the design effect.

An article was considered deficient on this criterion if between-cluster variation was not taken into account in calculating the sample size or in discussing the power of the trial.

Between-Cluster Variation Accounted for in the Analysis

From the preceding, it can be seen that if the between-cluster variation is not taken into account in the analysis, the standard error of the estimated intervention effect will be too small, so the confidence interval will be too narrow. The corresponding P value will also be too low, possibly leading to a spuriously statistically significant test result. For example, Donner and Klar[26] show that, for the data of Murray et al.,[20] the P value would be .03 if the effect of clustering were ignored, whereas it was greater than .1 after adjusting for the effect of clustering. Even though the estimated intracluster correlation coefficient was only .01, because there were approximately 120 subjects per cluster, the between-cluster variability more than doubled the variance of the observed treatment effect (i.e., the design effect was about 2.5).

Clustering must therefore be taken into account in the analysis. One way is to use the cluster as the unit of analysis: the outcome variable is then a summary statistic for all individuals in each cluster (e.g., mean change in body mass index, proportion of people smoking). Standard statistical methods can then be used to compare the cluster responses between different interventions, controlling for cluster-level baseline risk factors (e.g., cluster size, urban/rural location) if necessary. A disadvantage of this approach is that direct adjustment for individual-level covariates is not possible.

The alternative is to use the individual as the unit of analysis, taking into account the correlation between responses of individuals in the same cluster. For continuous outcome variables that are normally distributed, a mixed-effects analysis of variance (or covariance) is appropriate, with clusters nested within intervention groups.[27] For dichotomous outcome variables, a variety of new techniques are available, ranging from adjusted chi-square[26] to beta-binomial models[28] and generalized estimating equations.[29] An advantage of this last approach is that it permits direct adjustment for individual-level covariates. Detailed comparisons of these methods are provided by Donner and Klar.[26]

An article was considered deficient on this criterion if between-cluster variation was not taken into account in the analysis of the primary outcome variable(s).

Results

Table 1 shows the characteristics of the 21 prevention trials, 3 of which tested a specific protective measure[5–7]; the rest were concerned with a variety of health promotion or education interventions. Twelve trials were completely randomized, and the other 9 used designs that helped to ensure a balance of some prognostic factors between intervention groups: 3 used stratified randomization, 2 used a matched design, and in 4 trials the possible randomizations were restricted

TABLE 1—Characteristics of the 21 Cluster Randomized Trials Reviewed

Reference	Preventive Measure	Design	Unit of Randomization (Average Cluster Size)	No. Clusters Per Group	Primary Response Variable(s)	Follow-Up Time
2	Smoking cessation clinic	Completely randomized	Orientation session (11)	33, 31, 33	Sustained abstinence rates	1 y
3	HIV risk behavior reduction	Completely randomized	Community (200)	1, 2	Prevalence of HIV risk behaviors	6 mo
4	Mass media smoking prevention campaigns	Restricted randomization	Standard metropolitan statistical area (200)	2, 2, 2, 4	Adolescent smoking behaviors	2 y
5	Domestic water filter	Completely randomized	Household (4)	299, 308	Annual incidence of gastrointestinal illness	14 mo
6	Giardia control strategies	Completely randomized	Day-care centers (30)	11, 11, 10	Giardia prevalence	6 mo
7	Vitamin A supplementation	Completely randomized	Village (64)	229, 221	Prevalence of cough, fever, diarrhea	1 y
8	AIDS education program	Matched	Period of admission (47)	4, 4, 4	Change in attitudes, beliefs, skills, and behavior	10–18 wks
9	Nutrition intervention	Stratified randomization	Work site (275)	8, 8	Change in fat and fiber intake	15 mo
10	Weight control and smoking cessation programs	Completely randomized	Work site (200)	16, 16	Change in mean body mass index and smoking prevalence	2 y
12	Cardiovascular health promotion	Completely randomized	Ambulance station (15)	10, 8, 6, 4	Change in cardiovascular risk factors	1 y
13	Tobacco use prevention program	Matched	School (170)	11, 11	Prevalence of tobacco use	3 y
14	Tobacco use prevention program	Stratified randomization	School (140)	8, 8, 8, 8, 16	Change in tobacco use prevalence	1 y
15	Breast self-examination training	Completely randomized	Community (350)	1, 1, 1, 1	Prevalence of regular breast self-examination	2 y
16	Drug use prevention program	Restricted randomization	School cohort year (210)	6, 6, 12	Drug use indices	1 y
17	Advice on diet, smoking, exercise	Completely randomized	Family (4)	673, 700	Change in cardiovascular risk factors	6 y
18	Drug use prevention program	Restricted randomization, factorial	School (250)	3, 3, 3, 3	Drug use indices	1 y
19	Cardiovascular health promotion	Completely randomized	School (220)	2, 2	Change in cardiovascular risk factors	6 mo
20	Tobacco use prevention program	Completely randomized	School unit (190)	12, 12, 12, 12	Tobacco use prevalence and incidence	2 y
21	Cardiovascular health promotion	Completely randomized	Family (1.5)	50, 47	Change in cardiovascular risk factors	1 y
22	Physician advice on exercise	Stratified randomization	Family medicine resident (17)	12, 12	Change in exercise duration and frequency	1 mo
23	Drug use prevention program	Restricted randomization	School (220)	10, 10, 10	Drug use prevalence, cognitive risk factors	2 y

to those that ensured a balance on one or more important factors.

The average cluster size shown is the average number of eligible individuals who were randomized as an intact group. This may be considerably less than the actual size of the unit of randomization; for example, a community of 50 000 people may be randomized, but only 200 individuals within that community may be eligible or randomly sampled. The average cluster size ranged from 1.5 teenagers per family to about 350 women per community; 12 trials had very large clusters of more than 100 individuals.

The number of clusters per group shown in Table 1 is the number that were randomized, although some may have been lost to follow-up and therefore omitted from the analysis. One trial used a 2 × 2 factorial design with three clusters per group so that there were six clusters per group for testing the main effect of each intervention.[18] Six of the trials had very few clusters (four or fewer) per intervention group, and two had only one cluster in one or more groups. There is a tendency for the number of clusters per group to decrease as the size of the cluster increases.

Table 1 also shows the variety of primary response variables, which were often dichotomous, particularly when use of tobacco or other drugs was being studied. Follow-up times ranged from 1 month to 6 years. In nine trials, the

primary outcome was measured as a within-subject change from the baseline value, thus controlling for initial differences between subjects.

In most prevention trials, subjects in a cohort are measured before and after the intervention; when there is high turnover or a large number of subjects per cluster, however, repeat cross-sectional samples may be taken instead or as well.[24,30] Of the 21 trials, 16 used cohort samples, 4 used both cohort and repeat cross-sectional samples,[4,6,10,14] and 1 took repeat cross-sectional samples with an estimated 70% overlap (responses were anonymous).[3]

Table 2 gives the results of our evaluation of the 21 trials according to the two methodological criteria. Only 5 trials mentioned sample size or power, and, of these, 4 took clustering into account in the calculations.

In the statistical analysis, 12 of the 21 trials took account of clustering. Of these, 5 used the cluster as the unit of analysis,[4,7,10,13,14] 3 used a mixed-effects analysis of variance or covariance with clusters nested within treatment groups,[3,12,20] 3 estimated the design effect and used it to adjust standard errors or chi-square statistics,[2,5,23] and 1 used pedigree analysis.[21]

Studies that do not take clustering into account in the analysis are likely to be misleading in their conclusions. Of the 16 trials that reported a significant effect of the intervention on at least one of the primary outcome variables, 7 did not take clustering into account in their analysis. The reported statistical significance in these studies may therefore be spurious. Of the 5 trials that did not report a significant effect, only 2 discussed sample size or power; the other 3 may have lacked the power to detect an important effect.

In planning a cluster randomized trial, it is necessary to have an a priori estimate of either the design effect or the intracluster correlation coefficient, ρ. It is therefore recommended that trial reports include such estimates.[31] Of the 21 articles included here, 2 reported the design effects explicitly,[2,23] 2 gave estimates of ρ,[20,21] and another 2 gave tables of results by cluster, from which ρ could be estimated.[9,10]

Discussion

It is clear from our results that the methodological issues associated with cluster randomization are not recognized widely enough. Fewer than 60% of research teams reporting the results of recent cluster randomized primary prevention trials took clustering into account in the analysis. This is true even though it has been more than 16 years since these issues were discussed at a symposium on community intervention trials, articles from which were published in the *American Journal of Epidemiology* in 1978.[32] There have been many subsequent papers highlighting these issues and describing appropriate methods of analysis in a variety of journals. A recent review of the use of cluster randomization in epidemiology was provided by Donner and Klar.[33]

Many researchers had particular difficulty with dichotomous outcome variables, either treating them as continuous and using analysis of variance or ignoring the clustering completely for these outcomes. Methods for analyzing dichotomous variables in cluster randomized trials need to be made more widely known and more accessible through standard computer software packages for statistical analysis.

The results of our survey were even more discouraging when it came to taking between-cluster variation into account in sample size calculations or discussions of power. In most cases, there was simply no mention of these issues. It is possible that, for some studies, a previously published article reported sample size calculations, resulting in our judgment being too harsh on this criterion. Nevertheless, in no case was reference made to any such calculation. We believe that this is such an essential step in the planning of any trial that the chosen sample size should always be justified in the report of the trial results, which should either give details of the calculations or refer explicitly to where such details can be found. For cluster randomized trials, consideration of power at the planning stage is especially important because the number of clusters is often limited by availability or feasibility. Many subjects, extended follow-up times, and extensive use of resources are often involved, so it is essential to determine beforehand whether such a trial is likely to be large enough to achieve its goal. The number of clusters, rather than the size of the clusters, is most important in determining power.

It is interesting to note that the results of this review are almost identical to those of a review by Donner et al. of cluster randomized nontherapeutic intervention trials over a 10-year period.[31] They found that only 3 of the 16 trials (19%) that they reviewed accounted for between-cluster variability in sample size or power calculations, the same proportion as our 4 of 21 (19%). They also found that 8 of the 16 trials (50%) took account of the effect of clustering in the analysis, similar to our 57% (12 of 21).[31] In their review of 8 school-based drug use prevention trials, Ennett et al. found that 6 (75%) did not take account of cluster randomization in the statistical analysis.[34]

If a trial has only one cluster per intervention group, it is not possible to analyze it properly because the effect of intervention is totally confounded with differences between clusters. This design should therefore be discouraged for prevention trials currently in the planning stage.

Because it is often not possible to randomize many clusters in primary prevention trials, the probability of imbalance between intervention groups on important prognostic factors may be quite high. It is therefore particularly important in cluster randomized trials to ensure that the analysis of the intervention effect is not confounded by differences between groups in baseline risk factors. Reports of cluster randomized trials should always include a table showing the baseline distribution of important characteristics by intervention group; this table should also provide the number of clusters and the average cluster size for each group, since the overall precision of a study depends directly on these two quantities (as shown in equation 1). Such baseline descriptions are also helpful in determining to whom results of a particular trial can be generalized. This is a more

TABLE 2—Classification of 21 Articles on Primary Prevention Trials according to Whether They Satisfied Two Methodological Criteria for Cluster Randomized Trials

Criterion	Yes	No
Between-cluster variation accounted for in sample size and/or power calculations	4	17[a]
Between-cluster variation accounted for in analysis	12	9[b]

[a]Of these, three found no significant effect, possibly as a result of lack of power.
[b]Of these, seven found a significant effect that may have been spurious.

complicated problem than in trials randomizing individuals because generalizability may be determined both by characteristics of individual participants (e.g., age, sex) and by characteristics of clusters (e.g., type of work site).

Imbalances in important prognostic factors between intervention groups should be controlled for during the analysis, either by including them as covariates or by measuring change from baseline as the outcome variable (or both, if there are both baseline outcomes and other risk factors to be controlled). Alternatively, it may be helpful to prestratify or match on important cluster-level prognostic factors at the design stage, because this can increase the power to detect clinically relevant effects of intervention.[35] However, this gain in power can be accomplished only if the baseline risk factors are associated with the outcome, if the stratification and matching are taken into account during analysis, and if the gain in precision offsets the loss in degrees of freedom arising from such adjustments.[36] It becomes increasingly difficult to fulfill all three criteria when there are few clusters in each intervention group. For this reason, researchers are urged to proceed cautiously when selecting an experimental design. These problems could probably be avoided if sufficient attention were paid to determining sample size when designing a trial. Details required to determine sample size for completely randomized, stratified, and matched trials have been provided by Hsieh,[37] Donner,[38] and Shipley et al.,[39] respectively. Only 5 of the 21 trials dealt with prognostic factors by prestratifying (3 trials) or matching (2 trials), and none of these trials took the stratification or matching into account in the analysis.

Loss to follow-up is also an important issue in cluster randomized trials. If, for example, there are problems with industrial relations at a particular work site, not only might it be impossible to deliver the assigned intervention, but the whole cluster may be lost to follow-up. In addition, in trials that are cluster randomized, the focus of the intervention is often the cluster, not the individual, so individuals may be more likely to drop out. Not only might there be natural attrition as individuals leave a school, workplace, or community, but immigration into a cluster may also be allowed in some trials. Authors should therefore report at least the discrepancy between the number of subjects entering the trial and the number analyzed for each intervention group; in 17 of the 21 trials reviewed, this was done. The more complete reports also included an analysis of whether the loss to follow-up was related to prognostic factors differentially between intervention groups.

Our results perhaps give too rosy an impression of the present situation in regard to primary prevention trials involving clusters of subjects: we rejected many articles because the trials were not randomized at all, but clusters were merely assigned to a particular intervention, often for political or logistical reasons (e.g., Perry et al.,[40] Resnicow et al.[41]). Many researchers now realize the weakness of this "quasi-experimental" design due to its potential for bias, so it is to be hoped that the use of such a design will diminish.

In many primary prevention trials, there is no alternative to cluster randomization (e.g., when the intervention is delivered by mass media or in school classrooms or when structural changes are made in the workplace). In others, it may not be necessary; for example, we found some family practice studies in which individuals had been randomized, say to a physician-delivered smoking intervention[42] or to physician counseling on children's use of bicycle helmets,[43] although many randomize by practice for convenience or practicality. Because cluster randomized trials have less power than individually randomized trials of the same size, researchers should always consider carefully whether cluster randomization is justified. □

Acknowledgments

Dr Klar's research was supported by an Ontario Ministry of Health postdoctoral fellowship. Dr Donner's research was supported by a grant from the Natural Sciences and Engineering Research Council.

This review was done while Dr Simpson was spending a sabbatical period in the Department of Epidemiology and Biostatistics, University of Western Ontario.

References

1. Mausner JS, Kramer S. *Epidemiology: An Introductory Text.* 2nd ed. Philadelphia, PA: WB Saunders Co; 1985.
2. Lando HA, McGovern PG, Barrios FX, Etringer BD. Comparative evaluation of American Cancer Society and American Lung Association smoking cessation clinics. *Am J Public Health.* 1990;80:554–559.
3. Kelly JA, St. Lawrence JS, Diaz YE, et al. HIV risk behavior reduction following intervention with key opinion leaders of population: an experimental analysis. *Am J Public Health.* 1991;81:168–171.
4. Bauman KE, LaPrelle J, Brown JD, Koch GC, Padgett CA. The influence of three mass media campaigns on variables related to adolescent cigarette smoking: results of a field experiment. *Am J Public Health.* 1991;81:597–604.
5. Payment P, Richardson L, Siemiatycki J, Dewar R, Edwardes M, Franco E. A randomized trial to evaluate the risk of gastrointestinal disease due to consumption of drinking water meeting current microbiological standards. *Am J Public Health.* 1991;81:703–708.
6. Bartlett AV, Englender SJ, Jarvis BA, Ludwig L, Carlson JF, Topping JP. Controlled trial of *Giardia lamblia:* control strategies in day care centers. *Am J Public Health.* 1991;81:1001–1006.
7. Abdeljaber MH, Monto AS, Tilden RL, Schork A, Tarwotjo I. The impact of vitamin A supplementation on morbidity: a randomized community intervention trial. *Am J Public Health.* 1991;81:1654–1656.
8. McCusker J, Stoddard AM, Zapka JG, Morrison CS, Zorn M, Lewis BF. AIDS education for drug abusers: evaluation of short-term effectiveness. *Am J Public Health.* 1992;82:533–540.
9. Sorensen G, Morris DM, Hunt MK, et al. Work-site nutrition intervention and employees' dietary habits: the Treatwell program. *Am J Public Health.* 1992;82:877–880.
10. Jeffery RW, Forster JL, French SA, et al. The Healthy Worker Project: a work-site intervention for weight control and smoking cessation. *Am J Public Health.* 1993;83:395–401.
11. Ellickson PL, Bell RM, McGuigan K. Preventing adolescent drug use: long-term results of a junior high program. *Am J Public Health.* 1993;83:856–861.
12. Gomel M, Oldenburg B, Simpson JM, Owen N. Work-site cardiovascular risk reduction: a randomized trial of health risk assessment, education, counseling, and incentives. *Am J Public Health.* 1993;83:1231–1238.
13. Elder JP, Wildey M, de Moor C, et al. The long-term prevention of tobacco use among junior high school students: classroom and telephone interventions. *Am J Public Health.* 1993;83:1239–1244.
14. Sussman S, Dent CW, Stacy AW, et al. Project Towards No Tobacco Use: 1-year behavior outcomes. *Am J Public Health.* 1993;83:1245–1250.
15. Worden JK, Solomon LJ, Flynn BS, et al. A community-wide program in breast self-examination training and maintenance. *Prev Med.* 1990;19:254–269.
16. Graham JW, Johnson CA, Hansen WB, Flay BR, Gee M. Drug use prevention programs, gender, and ethnicity: evaluation of three seventh-grade Project SMART cohorts. *Prev Med.* 1990;19:305–313.
17. Knutsen SF, Knutsen R. The Tromso survey: the family intervention study—the effect of intervention on some coronary risk factors and dietary habits, a 6-year follow-up. *Prev Med.* 1991;20:197–212.
18. Hansen WB, Graham JW. Preventing alcohol, marijuana, and cigarette use among adolescents: peer pressure resistance training versus establishing conservative norms. *Prev Med.* 1991;20:414–430.
19. Arbeit ML, Johnson CC, Mott DS, et al. The Heart Smart cardiovascular school health promotion: behavior correlates of

risk factor change. *Prev Med.* 1992;21:18–32.
20. Murray DM, Perry CL, Griffin G, et al. Results from a statewide approach to adolescent tobacco use prevention. *Prev Med.* 1992;21:449–472.
21. Walker RW, Heller R, Redman S, O'Connell D, Boulton J. Reduction of ischemic heart disease risk markers in the teenage children of heart attack patients. *Prev Med.* 1992;21:616–629.
22. Lewis BS, Lynch WD. The effect of physician advice on exercise behavior. *Prev Med.* 1993;22:110–121.
23. Bell RM, Ellickson PL, Harrison ER. Do drug prevention effects persist into high school? How Project ALERT did with ninth graders. *Prev Med.* 1993;22:463–483.
24. Koepsell TD, Wagner EH, Cheadle AC, et al. Selected methodological issues in evaluating community-based health promotion and disease prevention programs. *Annu Rev Public Health.* 1992;13:31–57.
25. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. *Am J Epidemiol.* 1981;114:906–914.
26. Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *Am J Epidemiol.* 1994;140:279–289.
27. Koepsell TD, Martin DC, Diehr PH, et al. Data analysis and sample size issues in evaluations of community-based health promotion and disease prevention programs: a mixed-model analysis of variance approach. *J Clin Epidemiol.* 1991;44:701–713.
28. Williams DA. The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics.* 1975;31:949–952.
29. Zeger SL, Liang K-Y, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics.* 1988;44:1049–1060.
30. Feldman HA, McKinlay SM. Cohort versus cross-sectional design in large field trials: precision, sample size, and a unifying model. *Stat Med.* 1994;13:61–78.
31. Donner A, Brown KS, Brasher P. A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979–1989. *Int J Epidemiol.* 1990;19:795–800.
32. Hulley SB. Symposium on CHD prevention trials: design issues in testing life style intervention: introduction. *Am J Epidemiol.* 1978;108:85–86.
33. Donner A, Klar N. Cluster randomization in epidemiology: theory and application. *J Stat Plann Inference.* 1994;42:37–56.
34. Ennett ST, Tobler NS, Ringwalt CL, Flewelling RL. How effective is drug abuse resistance education? A meta-analysis of Project DARE outcome evaluations. *Am J Public Health.* 1994;84:1394–1401.
35. Freedman LS, Green SB, Byar DP. Assessing the gain in efficiency due to matching in a community intervention study. *Stat Med.* 1990;9:943–952.
36. Martin DC, Diehr P, Perrin EB, Koepsell TD. The effect of matching on the power of randomized community intervention studies. *Stat Med.* 1993;12:329–338.
37. Hsieh FY. Sample size formulae for intervention studies with the cluster as unit of randomization. *Stat Med.* 1988;8:1195–1201.
38. Donner A. Sample size requirements for stratified cluster randomization designs. *Stat Med.* 1992;11:743–750.
39. Shipley MJ, Smith PG, Dramaix M. Calculation of power for matched pair studies when randomization is by group. *Int J Epidemiol.* 1989;18:457–461.
40. Perry CL, Kelder SH, Murray DM, Klepp K-I. Communitywide smoking prevention: long-term outcomes of the Minnesota Heart Health Program and the Class of 1989 Study. *Am J Public Health.* 1992;82:1210–1216.
41. Resnicow K, Cross D, Lacosse J, Nichols P. Evaluation of a school-site cardiovascular risk factor screening intervention. *Prev Med.* 1993;22:838–856.
42. Hebert JR, Kristeller J, Ockene JK, et al. Patient characteristics and the effect of three physician-delivered smoking interventions. *Prev Med.* 1992;21:557–573.
43. Cushman R, James W, Waclawik H. Physicians promoting bicycle helmets for children: a randomized trial. *Am J Public Health.* 1991;81:1044–1046.