

# Design and Analysis of Group-Randomized Trials: A Review of Recent Practices

We reviewed group-randomized trials (GRTs) published in the *American Journal of Public Health and Preventive Medicine* from 1998 through 2002 and estimated the proportion of GRTs that employ appropriate methods for design and analysis.

Of 60 articles, 9 (15.0%) reported evidence of using appropriate methods for sample size estimation. Of 59 articles in the analytic review, 27 (45.8%) reported at least 1 inappropriate analysis and 12 (20.3%) reported only inappropriate analyses. Nineteen (32.2%) reported analyses at an individual or subgroup level, ignoring group, or included group as a fixed effect.

Hence increased vigilance is needed to ensure that appropriate methods for GRTs are employed and that results based on inappropriate methods are not published. (*Am J Public Health*. 2004;94:393–399)

| Sherri P. Varnell, PhD, David M. Murray, PhD, Jessica B. Janega, MS, and Jonathan L. Blitstein, MS

**DURING THE PAST 25 YEARS,** increased attention has been devoted to exploring the impact of intraclass correlation (ICC) in the design and analysis of group-randomized trials (GRTs) and to identifying appropriate methods for these trials. Despite this attention, periodic reviews of published GRTs have found that many investigators employed methods that do not account for the ICC properly.

A 1990 review of GRTs published in medical and epidemiological journals between 1979 and 1989<sup>1</sup> found that only 3 (19%) of the 16 reviewed articles accounted for ICC properly in sample size calculations, and only 8 (50%) accounted for ICC in the analysis. A meta-analysis of evaluations of 8 separate trials of a school-based program to

prevent drug use reported that only 2 (25%) accounted for ICC in the analysis.<sup>2</sup> Simpson et al.<sup>3</sup> reviewed all GRTs published in the *American Journal of Public Health and Preventive Medicine* between 1990 and 1993; they reported that only 4 (19%) of 21 articles included power calculations and only 12 (57%) included analyses that took ICC into account. A more recent review of community health interventions<sup>4</sup> included 8 GRTs; only 1 (12%) reported taking ICC into account properly in sample size calculations, though 7 (88%) accounted for ICC in the analysis.

In the meantime, methodologists have continued to focus attention on valid methods for estimating sample size and analyzing data from GRTs; a summary of

the work published in the last 5 years is provided in another article in this issue.<sup>5</sup> However, no recent review of published GRTs has examined the effect of this increased attention on the practices of investigators who conduct GRTs.

It is important to continue to monitor the published literature to determine the impact of recent methodological developments. Such reviews enable methodologists to determine the extent to which issues of clustering are recognized among investigators and to identify areas that may need further attention. They also alert investigators to attend more closely to the issues that they are missing. The goals of this study were to review GRTs recently published in the *American Journal of Public Health* and *Preven-*

*tive Medicine*, determine the extent to which authors provided evidence of taking ICC into account properly in the design and analysis, and compare our results to prior reviews.

## METHODS

We searched issues of the *American Journal of Public Health* and *Preventive Medicine* published from January 1998 to December 2002, inclusive, selecting all articles reporting the results of GRTs. These 2 journals were chosen for review to make possible a direct comparison to the review by Simpson et al.<sup>3</sup> GRTs were defined as studies that randomized intact social groups to study conditions but obtained observations from individuals nested within groups; we use the term *group* to designate the unit of assignment and *condition* to designate the experimental condition to which the group is assigned. Articles reporting the results of studies in which groups were not randomly assigned to study conditions were excluded, as were studies involving only observations at the group, rather than the individual, level, because these studies do not involve group-level ICC. We also excluded articles that did not include a clear statement indicating that all groups were randomized to study conditions, as well as articles indicating that some groups were randomized whereas others were nonrandomly assigned to conditions.

Each article was reviewed by the first 2 authors and by at least 1 of the other 2 authors to determine whether the article included sample size calculations and analyses taking ICC into account properly. In addition, because the design of a large GRT is often de-

scribed in detail in a background article that is cited in subsequent publications, we reviewed any references to sample size in background papers cited in the articles.

We reviewed articles to determine whether authors included evidence of taking clustering into account in arriving at the number of groups assigned to study conditions, such as the expected ICC, the group component of variance, or the variance inflation factor (VIF).<sup>6</sup> If no such evidence existed, articles were reviewed to determine whether the authors claimed that variance was inflated to account for the expected ICC, even if they provided no details on how it was done.

Because many articles presented more than 1 analytic strategy, we reviewed each article to determine whether all the analytic approaches used to evaluate intervention effects were appropriate, some were appropriate, or none were appropriate for GRTs. Also, each of the analytic approaches reported by the authors was recorded, along with any justifications offered by authors who reported inappropriate analytic strategies. Disagreements among reviewers on sample size or analytic ratings were resolved through roundtable discussion.

Murray<sup>7</sup> and Donner and Klar<sup>8</sup> have provided an extensive review of analytic methods appropriate for GRTs; Murray et al.<sup>5</sup> have provided a review of even more recent analytic developments. Table 1 presents the criteria used to judge whether analytic approaches reported in each article were appropriate. Methods considered appropriate for GRTs included but were not limited to mixed-model regression approaches including analysis of variance/analysis of covariance (ANOVA/

ANCOVA) and random coefficients models, 2-stage analyses (analysis on a summary statistic computed at the level of the group including randomization-based tests), and generalized estimating equations. Because each of these methods may be applied incorrectly, we established additional criteria for rating analyses as appropriately applied; these depend on the design of the study, the assumptions underlying the analytic method, and the robustness of the method to violations of these assumptions. Mixed-model analysis of variance/analysis of covariance (ANOVA/ANCOVAs) were considered appropriate if variation at the condition level was assessed against variation at the group level, with degrees of freedom (*df*) based on the number of groups, and with 1 or 2 time points included in the analysis. If more than 2 time points are included in the analysis of a GRT, a random-coefficients analysis preserves the nominal type I error rate, whereas mixed-model ANOVA/ANCOVAs may not<sup>9</sup>; thus, mixed-model ANOVA/ANCOVAs were considered inappropriate for GRTs with more than 2 time points whereas random-coefficient analyses were considered appropriate. Two-stage approaches were considered appropriate if the second stage was conducted at the group level with *df* based on the number of groups. A generalized estimating equations approach was considered appropriate if the analysis included 40 groups or more, because the type I error rate is unreliable with fewer groups.<sup>9,10</sup> Several articles reporting less common analytic methods referenced papers outlining these methods; we reviewed those articles for

evidence that the analytic method described was suitable for the analysis of GRTs.

## STUDY CHARACTERISTICS

A total of 60 articles that met the inclusion criteria were identified in the *American Journal of Public Health* and *Preventive Medicine* during the period January 1998 to December 2002, inclusive.<sup>11–70</sup> We also identified and reviewed 27 background papers referred to in the sample of reviewed articles.<sup>16,30,71–95</sup> Twenty-seven (45.0%) of the 60 reviewed articles were published in the *American Journal of Public Health*, and 33 (55.0%) were published in *Preventive Medicine*. Table 2 presents design characteristics of the studies described in the articles. The number of GRTs published per year in these journals more than doubled since the earlier review by Simpson et al.,<sup>3</sup> from 5.3 per year to 12 per year.

## SAMPLE SIZE CALCULATIONS

Overall, only 9 (15.0%) of the 60 papers reported an ICC, group component of variance, or VIF used for estimating the sample size for the trial, either in the reviewed article or in a background paper. Authors of an additional 3 (5.0%) articles claimed, either in the reviewed article or in a background paper, that variance had been inflated to account for the expected ICC but provided no evidence such as an ICC, variance components, or VIF.

## ANALYSIS

We excluded 1 article published in *Preventive Medicine* from

**TABLE 1—Common Analysis Options for Group-Randomized Trials and Conditions for Appropriate Application**

Method	Appropriate Application in GRTs
Mixed-model methods	
Repeated measures ANOVA/ANCOVA	1 or 2 time points
Random-coefficients approach	> 2 time points
Generalized estimating equations	
With small-sample correction	< 40 groups included in analysis
With no correction	≥ 40 groups included in analysis
2-stage methods (analysis on group means or other summary statistic)	Applied at level of unit of assignment
Post-hoc correction based on external estimates of ICC	Validity depends on validity of external estimates
Analysis at subgroup level, ignoring group-level ICC	Not appropriate for GRTs
Analysis at individual level, ignoring group-level ICC	Not appropriate for GRTs

Note. GRT = group-randomized trial; ANOVA = analysis of variance; ANCOVA = analysis of covariance; ICC = intraclass correlation.

the analytic review because the authors did not provide enough detail to determine whether the analytic strategy was appropriate. Table 3 presents the number and percentage of the 59 remaining articles that reported only appropriate methods; some appropriate and some inappropriate methods; or only inappropriate methods. Note that percentages within subsections of Table 3 may not add to the subsection total, because the categories were not mutually exclusive.

Among the 27 articles selected from the *American Journal of Public Health*, 18 (66.7%) reported only analyses taking ICC into account properly, 5 (18.5%) reported some analyses that took ICC into account properly and some that did not, and 4 (14.8%) reported only analyses that did not take ICC into account properly. The 32 *Preventive Medicine* articles reviewed included 14 (43.8%) that reported only analyses that took ICC into account properly, 10 (31.3%) that included some analyses that took ICC into account properly and some that did not, and 8 (25.0%) that reported only

analyses that did not take ICC into account properly.

Of the 15 articles that reported a mix of appropriate and inappropriate methods, 2 reported methods judged to be inappropriate that did not fall into any of the previously identified categories. One article reported a test of the intervention effect in a stratified analysis with an error term based on group variance rather than variance attributable to the interaction of the group and the stratum. Another article reported an analysis using *df* based on the number of groups for tests of the main effects, but used individual *df* for interaction terms.

Nineteen papers (32.2%) reported using analytic methods that ignored the group entirely, that included the group as a fixed effect, or that were conducted at a subgroup level. Among the 27 articles we reviewed from the *American Journal of Public Health*, 7 papers (25.9%) reported 1 of these analytic methods, whereas the corresponding figure among the 32 articles reviewed from *Preventive Medicine* was 12 (37.5%).

## DISCUSSION

It has been 25 years since Cornfield first drew attention in the public health literature to the unique design and analytic issues presented by GRTs.<sup>96</sup> Since then, an ever-increasing number of papers have appeared presenting comparisons of analytic methods, methods for sample size estimation, and reviews of published GRTs. Recent books on the design and analysis of GRTs have offered a comprehensive overview of relevant methods.<sup>7,8</sup> At the same time, analysis software that accommodates data structures common to GRTs has become more readily available and more user-friendly. Thus, many of the barriers that in the past may have prevented investigators from properly designing and analyzing GRTs have been removed. The purpose of this review was to determine the extent to which investigators have employed appropriate design and analytic strategies in recent publications, to describe the appropriate and inappropriate strategies used, and to compare the results with those of previous reviews.

The results of our review of methods used for sample size estimation were discouraging. In the entire sample of studies, less than one fifth provided clear evidence of sample size estimation taking ICC into account. Although it is possible that many of these studies performed such power calculations but did not report them, 27 (45.8%) of the reviewed studies had fewer than 10 groups per condition. Of the 27, only 1 reported evidence of estimating sample size using methods appropriate for GRTs. Thus, it is also likely that many investigators reporting small GRTs planned their study without considering issues of sample size carefully.

The proportion of articles reporting an ICC, a group component of variance, or the VIF (15.0%) was lower than that reported by Simpson et al.,<sup>3</sup> who found that 6 (29%) of their 21 articles reviewed reported that information. It is important that investigators who design and analyze GRTs make available the ICC estimates from their studies because it continues to be difficult for investigators to find estimates of ICCs to use in the planning stages of studies. The reporting of ICC estimates, whether in a main results paper or in a separate design or methods paper, should be common practice. Many such articles that focus on reporting ICC estimates and offering unadjusted and adjusted estimates have been published (see reference 5), but as our review shows, a large proportion of the investigators who could make estimates available do not.

The results of the review of analytic methods were mixed. The percentage of articles reporting only appropriate analytic methods (54.2%) was slightly lower than the 57% found in the review by Simpson et al.<sup>3</sup> How-

**TABLE 2—Characteristics of Studies Described in the 60 Reviewed Articles**

Characteristic	Articles	
	No.	%
Number of study conditions		
2	51	85.0
3	6	10.0
≥ 4	3	5.0
Matching or stratification in design		
Matching	22	36.7
Stratification	18	30.0
Matching and stratification	7	11.7
Randomization without matching or stratification	13	21.7
Type of group		
Schools or colleges	17	28.3
Worksites	11	18.3
Medical practices	9	15.0
Communities, neighborhoods, or postal networks	9	15.0
Housing projects or apartment buildings	3	5.0
Churches	3	5.0
Other	8	13.3
Number of groups per condition		
1 group	3	5.0
2-3 groups	5	8.3
4-5 groups	7	11.7
6-12 groups	20	33.3
13-25 groups	18	30.0
> 25 groups	7	11.7
Number of members per group		
< 10 members	8	13.3
10-50 members	19	31.7
51-100 members	16	26.7
> 100 members	17	28.3
Number of time points		
1 time point	2	3.3
2 time points	34	56.7
3 time points	17	28.3
4-9 time points	6	10.0
Number of time points varies within study	1	1.7
Design		
Cohort	38	63.3
Cross-sectional	13	21.7
Combination of cohort and cross-sectional	9	15.0
Primary outcome variables		
Smoking prevention or cessation	17	28.3
Dietary variables	12	20.0
Health screening	7	11.7
Alcohol, drug, or combination of alcohol, tobacco, drugs	5	8.3
Multiple health measures	5	8.3
Sun protection	3	5.0
Preventing physical or sexual abuse	2	3.3
Physician preventive practices	2	3.3
Workplace health and safety measures	2	3.3
Other	5	8.3

**TABLE 3—Results of the Review of Analytic Methods**

Criteria	No. (%) (n = 59)
Articles reporting only appropriate methods	32 (54.2)
Method	
Mixed-model methods with baseline measurement as covariate	10 (16.9)
Mixed-model ANOVA/ANCOVA approach with 1 or 2 time points	9 (15.3)
GEEs with ≥ 40 groups	2 (3.4)
2-stage analysis (analysis of group means or other summary statistics)	12 (20.0)
Articles reporting some appropriate and some inappropriate methods	15 (25.4)
Appropriate methods	
Mixed-model methods with baseline measurement as covariate	6 (10.2)
Mixed-model ANOVA/ANCOVA approach with 1 or 2 time points	4 (6.8)
GEEs with ≥ 40 groups	2 (3.4)
2-stage analysis	3 (5.1)
Inappropriate methods	
Analysis at an individual level, ignoring group-level ICC	9 (15.3)
Analysis at a subgroup level, ignoring group-level ICC	1 (1.7)
GEEs or other asymptotically robust method with < 40 groups	3 (5.1)
Other	2 (3.4)
Articles reporting only inappropriate methods	12 (20.3)
Method	
Analysis at an individual level, ignoring group-level ICC	6 (10.2)
Analysis at a subgroup level, ignoring group-level ICC	3 (5.1)
Analysis with group as a fixed effect	1 (1.7)
Mixed-model ANOVA/ANCOVA approach with > 2 time points	1 (1.7)
GEEs with < 40 groups	5 (8.5)

Note. GEE = generalized estimating equation; ANOVA = analysis of variance; ANCOVA = analysis of covariance; ICC = intraclass correlation.

ever, this comparison is not entirely fair because our review employed more stringent criteria for evaluating analytic methods as appropriate, based on the results of methodological work published after the review by Simpson et al.<sup>3</sup> In particular, we evaluated as inappropriate the use of generalized estimating equations and other asymptotically robust methods with few groups per condition and mixed-model repeated-measures analysis of variance with more than 2 time points, whereas Simpson et al. did not. Of the 59 papers included in the analytic review, 40 (67.8%) avoided methods con-

sidered inappropriate at the time of the review by Simpson et al., suggesting that awareness of the problems associated with those methods improved over the last 10 years. At the same time, we were surprised to find that 19 (32.2%) of the studies we reviewed reported analyses considered inappropriate 10 years ago. Clearly, some investigators, reviewers, and journal editors have still not heard or accepted the long-standing warnings against analysis at an individual or subgroup level that ignores the group-level ICC, or an analysis that includes the group as a fixed effect. And certainly some

investigators, reviewers, and journal editors are not familiar with the more recent developments that have identified other inappropriate methods. One method for ensuring a more careful screening of the design and analytic methods reported in published GRTs would be to require a review for the journal by a statistician or methodologist familiar with the unique analytic challenges presented by GRTs. Such measures are clearly needed, as reviews of appropriate methods and of published GRTs have not resulted in substantial change since the review by Simpson et al.<sup>3</sup>

The articles in our review that employed inappropriate analytic strategies generally did so without mentioning issues of clustering, although some did acknowledge clustering issues and offered justifications for ignoring them. One article that reported appropriate methods as well as inappropriate methods stated that the analyses ignoring group were performed because “individual variation in biological attributes could obscure important clinical changes in the school-level analyses as a result of reduced statistical power.”<sup>15</sup> We don’t disagree with the statement, but we would disagree with its use as justification for an analysis that is likely to carry an inflated type I error rate. The 2 “penalties” of GRTs, variance inflation and limited *df*, simply cannot be avoided.<sup>96,97</sup> Investigators should consider in advance whether the number of groups is adequate to permit the comparisons of interest, and if it is not, they should adopt methods to improve precision, likely to include adding more groups.

Another article that reported both appropriate and inappro-

priate methods gave as their justification that ICCs such as 0.024, 0.026, and 0.083 were “very small.”<sup>50(p158)</sup> This claim ignores the fact that variance inflation in a GRT depends on both the ICC and the average number of members per group.<sup>96</sup> Ignoring an ICC as small as 0.001 can be dangerous if the number of members per group is large; similarly, even with as few as 10 members per group, an ICC of 0.08 can inflate type I error rates beyond the nominal level.

One paper reported an appropriate analysis that yielded a negative estimate for the ICC for the outcome variable of interest. The authors concluded that because the group component of variance was estimated as negative, the “true” value of the group component should be zero, and they then reported analyses that ignored the group, with *df* calculated based on the number of members. Intuitively, this idea makes sense, because variances are squared quantities and cannot theoretically be negative. However, if the true value is zero and estimates are normally distributed around the true value, the estimate would be expected to be negative approximately half the time. Some argue that negative estimates should be set to zero, but simulation studies have demonstrated that this practice depresses power.<sup>9,97,98</sup> A negative variance component will result in a negative ICC and a VIF that is less than 1. The variance for the appropriate analysis will actually be larger, then, if variance components are constrained to be zero than if they are allowed to be estimated as negative. We considered the choice to set the group component of variance to

zero to be a conservative strategy but judged the analysis overall to be inappropriate because the investigators used *df* based on the number of individuals. This choice is appropriate when individuals are randomized to study conditions, but in this case, identifiable social groups were assigned to study conditions, and *df* should be based on the number of groups.<sup>96</sup> That distinction will not materially affect the outcome when there are a large number of groups assigned to each condition so that *df* based on groups are large, but it will if the number of groups per condition is small.

Three articles in our sample described studies that randomized only 1 group to each study condition. This design presents an analytic challenge without a satisfactory solution in that the group is confounded with the condition; thus, a proper test of the intervention effect is not possible. The authors of these articles offered no justification for employing such a problematic design.

## SUMMARY

The number of published GRTs continues to increase, as evidenced by our identification of 60 articles in 5 years, compared with the 21 articles found in 4 years in the review by Simpson et al.<sup>3</sup> Given this increase, and the expense and effort required to design and implement GRTs, it is imperative that investigators employ appropriate design and analytic methods. It is also imperative that reviewers and journal editors do a better job of screening submissions to prevent publication of GRTs that employ inappropriate analytic methods. ■

## About the Authors

Sherril P. Varnell is with Northrop-Grumman Mission Systems, Atlanta, Ga. David M. Murray, Jessica B. Janega, and Jonathan L. Blitstein are with the Department of Psychology, College of Arts and Sciences, University of Memphis, Tenn.

Requests for reprints should be sent to David M. Murray, PhD, 3693 Norriswood, 202 Psychology Building, Memphis, TN 38134 (e-mail: d.murray@mail.psy.memphis.edu).

This article was accepted September 18, 2003.

## Contributors

S.P. Varnell wrote the article. All authors reviewed the article, contributed to the results, and edited several versions of the article.

## Human Participant Protection

No protocol approval was needed for this study.

## References

1. Donner A, Brown KS, Brasher P. A methodologic review of non-therapeutic intervention trials employing cluster randomization, 1979–1989. *Int J Epidemiol.* 1990;19:795–800.
2. 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.
3. Simpson JM, Klar N, Donner A. Accounting for cluster randomization: a review of primary prevention trials, 1990 through 1993. *Am J Public Health.* 1995;85:1378–1383.
4. Smith PJ, Moffatt MEK, Gelskey SC, Hudson S, Kaita K. Are community health interventions evaluated appropriately? A review of six journals. *J Clin Epidemiol.* 1997;50:137–146.
5. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. *Am J Public Health.* 2004;94:423–432.
6. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. *Am J Epidemiol.* 1981;114:906–914.
7. Murray DM. *Design and Analysis of Group-Randomized Trials.* New York, NY: Oxford University Press; 1998.
8. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research.* London, England: Arnold; 2000.

9. Murray DM, Hannan PJ, Wolfinger RD, Baker WL, Dwyer JH. Analysis of data from group-randomized trials with repeat observations on the same groups. *Stat Med.* 1998;17:1581–1600.
10. Feng Z, McLerran D, Grizzle J. A comparison of statistical methods for clustered data analysis with Gaussian error. *Stat Med.* 1996;15:1793–1806.
11. Steptoe A, Kerry SM, Rink E, Hilton S. The impact of behavioral counseling on stage of change in fat intake, physical activity, and cigarette smoking in adults at increased risk of coronary heart disease. *Am J Public Health.* 2001;91:165–169.
12. Allen JD, Stoddard AM, Mays J, Sorensen G. Promoting breast and cervical cancer screening at the workplace: results from the Woman to Woman Study. *Am J Public Health.* 2001;91:584–590.
13. Rotheram-Borus MJ, Lee MB, Gwadz M, Drainin B. An intervention for parents with AIDS and their adolescent children. *Am J Public Health.* 2001;91:1294–1302.
14. Resnicow K, Jackson A, Wang T, et al. A motivational interviewing intervention to increase fruit and vegetable intake through black churches: results of the Eat for Life trial. *Am J Public Health.* 2001;91:1686–1693.
15. Hennrikus DJ, Jeffery RW, Lando HA, et al. The SUCCESS project: the effect of program format and incentives on participation and cessation in worksite smoking cessation programs. *Am J Public Health.* 2002;92:274–279.
16. Lazovich D, Parker DL, Brosseau LM, et al. Effectiveness of a worksite intervention to reduce an occupational exposure: the Minnesota Wood Dust Study. *Am J Public Health.* 2002;92:1498–1505.
17. Foshee VA, Bauman KE, Greene WF, Koch GG, Linder GF, MacDougall JE. The Safe Dates program: 1-year follow-up results. *Am J Public Health.* 2000;90:1619–1622.
18. Duan N, Fox SA, Derose KP, Carson S. Maintaining mammography adherence through telephone counseling in a church-based trial. *Am J Public Health.* 2000;90:1468–1471.
19. Cummings KM, Hyland A, Saunders-Martin T, Perla J, Coppola PR, Pechacek TF. Evaluation of an enforcement program to reduce tobacco sales to minors. *Am J Public Health.* 1998;88:932–935.
20. Mayer JA, Eckhardt L, Stepanski BM, et al. Promoting skin cancer prevention counseling by pharmacists. *Am J Public Health.* 1998;88:1096–1099.
21. Farrell KP, Brophy MC, Chisolm JJ, Rohde CA, Strauss WJ. Soil lead abatement and children's blood lead levels in an urban setting. *Am J Public Health.* 1998;88:1837–1839.
22. Noland MP, Kryscio RJ, Hinkle J, et al. Relationship of personal tobacco-raising, parental smoking, and other factors to tobacco use among adolescents living in a tobacco-producing region. *Addict Behav.* 1996;21:349–361.
23. Hillman AL, Ripley K, Goldfarb N, Naumah I, Weiner J, Lusk E. Physician financial incentives and feedback: failure to increase cancer screening in Medicaid managed care. *Am J Public Health.* 1998;88:1699–1701.
24. Chou C-P, Montgomery S, Pentz MA, et al. Effects of a community-based prevention program on decreasing drug use in high-risk adolescents. *Am J Public Health.* 1998;88:944–948.
25. Kellam SG, Anthony JC. Targeting early antecedents to prevent tobacco smoking: findings from an epidemiologically based randomized field trial. *Am J Public Health.* 1998;88:1490–1495.
26. Campbell MK, Demark-Wahnefried W, Symons M, et al. Fruit and vegetable consumption and prevention of cancer: the Black Churches United for Better Health project. *Am J Public Health.* 1999;89:1390–1396.
27. Walsh MM, Hilton J, Masouedis CM, Gee L, Chesney MA, Ernster VL. Smokeless tobacco cessation intervention for college athletes: results after 1 year. *Am J Public Health.* 1999;89:228–234.
28. Sorensen G, Stoddard A, Peterson K, et al. Increasing fruit and vegetable consumption through worksites and families in the Treatwell 5-a-Day study. *Am J Public Health.* 1999;89:54–60.
29. Harrell JS, McMurray RG, Gansky SA, Bangdiwala SI, Bradley CB. A public health vs a risk-based intervention to improve cardiovascular health in elementary school children: the Cardiovascular Health in Children Study. *Am J Public Health.* 1999;89:1529–1535.
30. Cameron R, Brown KS, Best A, et al. Effectiveness of a social influences smoking prevention program as a function of provider type, training method, and school risk. *Am J Public Health.* 1999;89:1827–1831.
31. Manfredi C, Crittenden KS, Cho YI, Engler J, Warnecke RB. The effect of a structured smoking cessation program, independent of exposure to existing interventions. *Am J Public Health.* 2000;90:751–756.
32. Perry CL, Bishop DB, Taylor G, et al. Changing fruit and vegetable consumption among children: the 5-A-Day Power Plus program in St Paul, Minnesota. *Am J Public Health.* 1998;88:603–609.
33. Wechsler H, Basch CE, Zybert P, Shea S. Promoting the selection of low-fat milk in elementary school cafeterias in an inner-city Latino community: evaluation of an intervention. *Am J Public Health.* 1998;88:427–433.
34. Resnicow K, Davis MA, Smith M, et al. Results of the TeachWell worksite wellness program. *Am J Public Health.* 1998;88:250–257.
35. Foshee VA, Bauman KE, Arriaga XB, Helms RW, Koch GG, Linder GF. An evaluation of Safe Dates, an adolescent dating violence prevention program. *Am J Public Health.* 1998;88:45–50.
36. Sorensen G, Stoddard A, Hunt MK, et al. The effects of a health promotion—health protection intervention on behavior change: the WellWorks Study. *Am J Public Health.* 1998;88:1685–1690.
37. Forster JL, Murray DM, Wolfson M, Blaine TM, Wagenaar AC, Hennrikus DJ. The effects of community policies to reduce youth access to tobacco. *Am J Public Health.* 1998;88:1193–1198.
38. Sussman S, Dent CW, Stacy AW, Craig S. One-year outcomes of Project Towards No Drug Abuse. *Prev Med.* 1998;27:632–642.
39. Gritz ER, Thompson B, Emmons K, Ockene JK, McLerran DF, Nielsen IR, the Working Well Research Group. Gender differences among smokers and quitters in the Working Well Trial. *Prev Med.* 1998;27:553–561.
40. Young JM, D'Este C, Ward JE. Improving family physicians' use of evidence-based smoking cessation strategies: a cluster randomization trial. *Prev Med.* 2002;35:572–583.
41. Brown KS, Cameron R, Madill C, et al. Outcome evaluation of a high school smoking reduction intervention based on extracurricular activities. *Prev Med.* 2002;35:506–510.
42. Ausems M, Mesters I, van Brueken G, De Vries H. Short-term effects of a randomized computer-based out-of-school smoking prevention trial aimed at elementary schoolchildren. *Prev Med.* 2002;34:581–589.
43. Zhu K, Hunter S, Bernard LJ, et al. An intervention study on screening for breast cancer among single African-American women aged 65 and older. *Prev Med.* 2002;34:536–545.
44. Segura J, Castells X, Casamitjana M, Macia F, Porta M, Katz SJ. A randomized controlled trial comparing three invitation strategies in a breast cancer screening program. *Prev Med.* 2001;33:325–332.
45. Aveyard P, Sherratt E, Almond J, et al. The change-in-stage and updated smoking status results from a cluster-randomized trial of smoking prevention and cessation using the transtheoretical model among British adolescents. *Prev Med.* 2001;33:313–324.
46. Sejr HS, Osler M. Do smoking and health education influence student nurses' knowledge, attitudes, and professional behavior? *Prev Med.* 2002;34:260–265.
47. Campbell MK, Tessaro I, DeVellis B, Benedict S, Kelsey K, Belton L. Effects of a tailored health promotion program for female blue-collar workers: Health Works for Women. *Prev Med.* 2002;34:313–323.
48. Baranowski T, Baranowski J, Cullen KW, et al. 5 a Day Achievement Badge for African-American Boy Scouts: pilot outcome results. *Prev Med.* 2002;34:353–363.
49. Dent C, Sussman S, Stacy AW. Project Towards No Drug Abuse: generalizability to a general high school sample. *Prev Med.* 2001;32:514–520.
50. Geller AC, Glanz K, Shigaki D, Isenc MR, Sun T, Maddock J. Impact of skin cancer prevention on outdoor aquatics staff: the Pool Cool Program in Hawaii and Massachusetts. *Prev Med.* 2001;33:155–161.
51. Mathey M-FM, Vanneste VGG, de Graff C, de Groot LC, van Stavener WA. Health effect of improved meal ambience in a Dutch nursing home: a 1-year intervention study. *Prev Med.* 2001;32:416–423.
52. Beresford SAA, Thompson B, Feng Z, Christianson A, McLerran D, Patrick DL. Seattle 5 a Day worksite program to increase fruit and vegetable consumption. *Prev Med.* 2001;32:230–238.
53. Hunt MK, Lobb R, Delichatsios H, Stone C, Emmons K, Gillman MW. Process evaluation of a clinical preventive nutrition intervention. *Prev Med.* 2001;33:82–90.
54. Delichatsios H, Hunt MK, Lobb R, Emmons K, Gillman MW. EatSmart: efficacy of a multifaceted preventive nutrition intervention in clinical practice. *Prev Med.* 2001;33:91–98.
55. Slater JS, Ha CN, Malone ME, et al. A randomized community trial to increase mammography utilization among low-income women living in public housing. *Prev Med.* 1998;27:862–870.
56. Solberg LI, Kottke TE, Brekke ML.

Will primary care clinics organize themselves to improve the delivery of preventive services? A randomized controlled trial. *Prev Med.* 1998;27:623–631.

57. Reynolds KD, Franklin FA, Binkley D, et al. Increasing the fruit and vegetable consumption of fourth-graders: Results from the High 5 Project. *Prev Med.* 2000;30:309–319.

58. Litrownik AJ, Elder JP, Campbell NR, et al. Evaluation of a tobacco and alcohol use prevention program for Hispanic migrant adolescents: promoting the protective factor of parent-child communication. *Prev Med.* 2000;31:124–133.

59. Dietrich AJ, Olson AL, Sox CH, Tosteson TD, Grant-Peterson J. Persistent increase in children's sun protection in a randomized controlled community trial. *Prev Med.* 2000;31:569–574.

60. Thompson NJ, Boyko EJ, Dominitz JA, et al. A randomized controlled trial of a clinic-based support staff intervention to increase the rate of fecal occult blood test ordering. *Prev Med.* 1999;30:244–251.

61. Richmond R, Kehoe L, Heather N, Wodak A. Evaluation of a workplace brief intervention for excessive alcohol consumption: the Workscreen Project. *Prev Med.* 2000;30:51–63.

62. Norris SL, Grothaus LC, Buchner DM, Pratt M. Effectiveness of physician-based assessment and counseling for exercise in a staff model HMO. *Prev Med.* 2000;30:513–523.

63. Manfredi C, Crittenden KS, Warnecke RB, Engler J, Cho YI, Shaligram C. Evaluation of a motivational smoking cessation intervention for women in public health clinics. *Prev Med.* 1999;28:51–60.

64. Tilley BC, Glanz K, Kristal AR, et al. Nutrition intervention for high-risk auto workers: results of the Next Step Trial. *Prev Med.* 1999;28:284–292.

65. Tilley BC, Vernon SW, Myers R, et al. The Next Step Trial: impact of a worksite colorectal cancer screening promotion program. *Prev Med.* 1999;28:276–283.

66. Lowensteyn I, Joseph L, Levinton C, Abrahamowicz M, Steinert Y, Grover S. Can computerized risk profiles help patients improve their coronary risk? The results of the Coronary Health Assessment Study (CHAS). *Prev Med.* 1998;27:730–737.

67. Basen-Engquist K, Hudmon KS, Tripp M, Chamberlain RM. Worksite health and safety climate: scale development and effects of a health promotion intervention. *Prev Med.* 1998;27:111–119.

68. Wu Z, Detels R, Zhang J, Li V, Li J. Community-based trial to prevent drug use among youths in Yunnan, China. *Am J Public Health.* 2002;92:1952–1957.

69. Hancock L, Sanson-Fisher RW, Perkins J, McClintock A, Howley P, Gibberd R. Effect of a community action program on adult quit smoking rates in rural Australian towns: the CART Project. *Prev Med.* 2001;32:118–127.

70. Hancock L, Sanson-Fisher RW, Perkins J, Giris A, Howley P, Schofield M. The effect of a community action intervention on adolescent smoking rates in rural Australian towns: the CART Project. *Prev Med.* 2001;32:332–340.

71. Allen JD, Sorensen G, Stoddard AM, Colditz GA, Peterson KE. Intention to have a mammogram in the future among women who have underused mammography in the past. *Health Educ Behav.* 1998;25:474–488.

72. Allen JD, Sorensen G, Stoddard AM, Peterson KE, Colditz GA. The relationship between social network characteristics and breast cancer screening practices among employed women. *Ann Behav Med.* 1999;21:193–200.

73. Hilton S, Doherty S, Kendrick T, et al. Promotion of healthy behaviour among adults at increased risk of coronary heart disease in general practice: methodology and baseline data from the Change of Heart Study. *Health Educ J.* 1999;58:3–16.

74. Foshee VA, Linder GF, Bauman KE, et al. The Safe Dates Project: theoretical basis, evaluation design, and selected baseline findings. *Am J Prev Med.* 1996;12(5 suppl):39–47.

75. Fox SA, Stein JA, Gonzalez RA, Farrenkopf M, Dellinger A. A trial to increase mammography utilization among Los Angeles Hispanic women. *J Health Care Poor Underserved.* 1998;9:309–321.

76. Pentz MA, Dwyer JH, MacKinnon DP, et al. A multicomponent trial for primary prevention of adolescent drug abuse. Effects on drug use prevalence. *JAMA.* 1989;261:3259–3266.

77. Kramish CM, Polhamus B, McClelland JW, et al. Assessing fruit and vegetable consumption in a 5 a day study targeting rural blacks: the issue of portion size. *J Am Diet Assoc.* 1996;96:1040–1042.

78. Campbell MK, Symons M, Demark-Wahnefried W, et al. Stages of change and psychosocial correlates of fruit and vegetable consumption among rural African-American church members. *Am J Health Promotion.* 1998;12:185–191.

79. Demark-Wahnefried W, McClelland JW, Campbell MK, et al. Aware-

ness of cancer-related programs and services among rural African Americans. *J Natl Med Assoc.* 1998;90:197–202.

80. McClelland JW, Demark-Wahnefried W, Mustian RD, Cowan AT, Campbell MK. Fruit and vegetable consumption of rural African Americans: baseline survey results of the Black Churches United for Better Health 5 A Day Project. *Nutr Cancer.* 1998;30:148–157.

81. Sorensen G, Hunt MK, Cohen N, et al. Worksite and family education for dietary change: the Treatwell 5-a-Day program. *Health Educ Res.* 1998;13:577–591.

82. Santi S, Best JA, Payne ME, Brown KS, Cameron R. A comparison between instructional experience and performance of teachers and nurses delivering a smoking prevention program. *Can J Public Health.* 1992;83:433–436.

83. Sorensen G, Himmelstein JS, Hunt MK, et al. A model for worksite cancer prevention: integration of health protection and health promotion in the Well-Works Project. *Am J Health Promotion.* 1995;10:55–62.

84. Forster JL, Wolfson M. Youth access to tobacco: policies and politics. *Annu Rev Public Health.* 1998;19:203–235.

85. Abrams DB, Boutwell WB, Grizzle J, Heimendinger J, Sorensen G, Varnes J. Cancer control at the workplace: the Working Well Trial. *Prev Med.* 1994;23:15–27.

86. Aveyard P, Cheng KK, Almond J, et al. Cluster randomised controlled trial of expert system based on the transtheoretical (“stages of change”) model for smoking prevention and cessation in schools. *BMJ.* 1999;319:948–953.

87. Cullen KW, Baranowski T, Baranowski J, et al. “5 A Day” achievement badge for urban Boy Scouts: formative evaluation results. *J Cancer Educ.* 1998;13:162–168.

88. Thompson B, van Leynseele J, Beresford S. Recruiting worksites to participate in a health promotion research study. *Am J Health Promotion.* 1997;11:344–351.

89. Thompson B, Shannon J, Beresford S, Jacobson PE, Ewing JA. Implementation aspects of the Seattle “5 a Day” intervention project: strategies to help employees make dietary changes. *Top Clin Nutr.* 1995;11:58–75.

90. Reynolds KD, Raczynski JM, Binkley D, et al. Design of “High 5”: a school-based study to promote fruit and vegetable consumption for reduction of cancer risk. *J Cancer Educ.* 1998;13:19–177.

91. Elder JP, Campbell NR, Litrownik AJ, et al. Predictors of cigarette and alcohol susceptibility and use among Hispanic migrant adolescents. *Prev Med.* 2000;31:115–123.

92. Dietrich AJ, Olson AL, Sox CH, et al. A community-based randomized trial encouraging sun protection for children. *Pediatrics.* 1998;102:E64.

93. Cebul RD. Randomized, controlled trials using the Metro Firm System. *Med Care.* 1991;29:JS9JS18.

94. Tilley BC, Vernon SW, Glanz K, et al. Worksite cancer screening and nutrition intervention for high-risk auto workers: design and baseline findings of the Next Step Trial. *Prev Med.* 1997;26:227–235.

95. Hancock L, Sanson-Fisher RW, Redman S, et al. Community action for cancer prevention: overview of the Cancer Action in Rural Towns (CART) project, Australia. *Health Promotion Int.* 1996;11:277–290.

96. Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol.* 1978;108:100–102.

97. Murray DM, Hannan PJ, Baker WL. A Monte Carlo study of alternative responses to intraclass correlation in community trials: is it ever possible to avoid Cornfield's penalties? *Evaluation Rev.* 1996;20:313–337.

98. Hannan PJ, Murray DM. Gauss or Bernoulli? A Monte Carlo comparison of the performance of the linear mixed model and the logistic mixed model analyses in simulated community trials with a dichotomous outcome variable at the individual level. *Evaluation Rev.* 1996;20:338–352.