Published in final edited form as:

Contemp Clin Trials. 2011 July ; 32(4): 477–484. doi:10.1016/j.cct.2011.03.001.

Cluster randomized trials of cancer screening interventions: are appropriate statistical methods being used?

Catherine M Crespi^{a,b,*}, Annette E. Maxwell^b, and Sheng Wu^a

^aDepartment of Biostatistics, University of California, Los Angeles, School of Public Health, Center for the Health Sciences 51-254, Box 951772, Los Angeles, CA 90095-1772

^bSchool of Public Health and Jonsson Comprehensive Cancer Center, University of California, Los Angeles, Center for the Health Sciences A2-125, Box 956900, Los Angeles, CA 90095-6900

Abstract

The design and analysis of cluster randomized trials can require more sophistication than individually randomized trials. However, the need for statistical methods that account for the clustered design has not always been appreciated, and past reviews have found widespread deficiencies in methodology and reporting. We reviewed cluster randomized trials of cancer screening interventions published in 1995–2010 to determine whether the use of appropriate statistical methods had increased over time. Literature searches yielded 50 articles reporting outcome analyses of cluster randomized trials of breast, cervix and colorectal cancer screening interventions. Of studies published in 1995–1999, 2000–2002, 2003–2006 and 2007–2010, 55% (6/11), 82% (9/11), 92% (12/13) and 60% (9/15) used appropriate analytic methods, respectively. Results were suggestive of a peak in 2003–2006 ($p=.06$) followed by a decline in 2007–2010 ($p=.08$). While the sample of studies was small, these results indicate that many cluster randomized trials of cancer screening interventions have had deficiencies in the application of correct statistical procedures for the outcome analysis, and that increased adoption of appropriate methods in the early and mid-2000's may not have been sustained.

Keywords

cluster randomized trials; intraclass correlation coefficient; CONSORT; cancer screening intervention trials

Introduction

In cluster randomized trials, natural groups or “clusters” of individuals such as families, health care providers or communities are randomized to intervention and control conditions. A cluster randomized design can have a number of advantages, such as reduced potential for contamination compared to an individually randomized trial and increased efficiency of intervention delivery [1–3]. In addition, many interventions are naturally applied at a group level (e.g., interventions directed at health care providers to promote cancer screening

© 2011 Elsevier Inc. All rights reserved.

*Corresponding author Phone: (310) 206-9364, Fax: (310) 206-3566, ccrespi@ucla.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

among patients) [2]. For these reasons, cluster randomized trials have become an increasingly important study design in intervention and clinical research [1, 3–5].

The design and analysis of cluster randomized trials can require more sophisticated approaches than individually randomized trials. Individuals who receive care from the same provider or live in the same community tend to be more alike than individuals who have different providers or live in different communities. As a result, in cluster randomized trials, the outcomes of individuals within the same cluster are statistically dependent rather than independent, a phenomenon that must be accounted for in the design and analysis of a trial [1–3, 6]. The degree of correlation among participants within the same cluster is typically quantified by the intraclass correlation coefficient (ICC). Several references provide comprehensive expositions of appropriate design and analysis methods for cluster randomized trials [1, 2, 7].

The special features of cluster randomized trials have not always been appreciated by investigators. Reviews of cluster randomized trials published in the 1990's and early 2000's found widespread deficiencies in conduct and reporting, with many studies not accounting for clustering in the planning stage, the data analysis or both, and not providing adequate reporting of study features [8–12]. In response to these deficiencies, the Consolidated Standards of Reporting Trials (CONSORT) statement extension to cluster randomized trials was published in March 2004 [13]. This statement extended the CONSORT statement [14, 15], which provides a framework for reporting parallel group individually randomized trials, to include reporting of the special features of cluster randomized trials, and provides a checklist of specific items to report.

While there are many features of a cluster randomized trial that should be accounted for and reported, perhaps of most importance is that the outcome analysis use appropriate statistical methodology that accounts for the clustered design. Failure to account for clustering generally results in an inflated Type I error rate (the probability of incorrectly rejecting the null hypothesis of no effect), leading to artificially small p-values and erroneous conclusions that a treatment is effective when it is not [2, 5, 16].

We identified cluster randomized trials of cancer screening interventions published from 1995 to 2010 and examined the methodology they used for the outcome analysis as a key quality indicator. Our research question was whether the use and reporting of appropriate statistical methods had increased over time, and in particular after the publication of the CONSORT extension to cluster randomized trials. We also examined whether the use of appropriate methods differed by journal impact, and whether studies reported ICCs for outcomes, which are useful for investigators planning future trials and are included in the CONSORT checklist of items to be reported.

Methods

We conducted a literature search to identify articles that reported outcome analyses of cluster randomized trials of cancer screening interventions. Our inclusion criteria for articles were: (1) receipt of breast, cervical or colon cancer screening as an outcome; (2) use of a cluster randomized design; (3) in English; and (4) published between January 1, 1995 and December 31, 2010. We used PubMed and Web of Science to identify papers that met our inclusion criteria. Due to the lack of standardized nomenclature for reporting the use of a cluster-randomized design, we searched broadly for papers reporting randomized trials of breast, cervical or colorectal cancer screening interventions, then screened the resulting collection of papers for those reporting cluster-randomized trials with a cancer screening outcome. In PubMed, searches were conducted utilizing the PubMed automatic MeSH

translation capability, with limits specified as (1) clinical trials, (2) English language, and (3) publication dates January 1, 1995 through December 31, 2010. Three searches were conducted using the terms “randomized trial [site] cancer screening,” where site was “colorectal,” “breast” or “cervical.” The searches were conducted on February 1, 2011. We supplemented these searches with citation searches of articles reporting study design, baseline data or preliminary results, in an attempt to find the associated outcome paper. We also presented our initial list of articles to cancer screening experts familiar with the literature to identify omitted studies.

Each article was independently reviewed by two statisticians to ascertain the statistical methodology used for the analysis of the cancer screening outcome variable. Authoritative texts [1, 2, 7] were used as a guide for classifying methods as appropriate or not. Since the style of reporting methods varied widely, identification of the statistical methods used usually necessitated an in-depth review of the Methods and Results sections of each article. Articles that made reference to an appropriate method were classified as using an appropriate method even if the article did not provide enough detail to conclusively determine that the method was applied appropriately. Initial disagreements between raters were settled by article re-review, discussion and consensus.

To assess whether rates of the use of appropriate methods varied by journal impact, we classified each article as “high influence” or “low influence” using Article Influence scores obtained from www.eigenfactor.org. This measure uses citation data to assess the influence of a journal in relation to other journals and is roughly analogous to the 5-Year Journal Impact Factor in that it is a ratio of a journal’s citation influence to the size of the journal’s article contribution over a period of five years. The mean score is 1.00. Articles with Article Influence scores greater than 1.00 were categorized as high influence; articles with scores less than 1.00 were categorized as low influence. Four articles did not have Article Influence scores; these articles were in publications with little or no coverage in citation databases and were classified as low influence.

We used symmetric nearest neighbor smoothing [17], implemented with the Stata user-contributed “running” command, to obtain smoothed estimates of the running proportion of studies using appropriate methods over time (i.e., proportions averaged over nearest neighbors in time). The binary indicator 0/1 for use of appropriate methods was logit transformed and a span of 0.4 was used. Logistic regression with linear and quadratic terms for year of publication was used to infer the statistical significance of overall time trends. Due to the relatively small sample sizes involved, we used exact methods to obtain estimates of the proportion of studies using appropriate methods by time interval and to test for differences in proportions between time intervals. For these analyses, we divided our time periods into four intervals, January 1, 1995–December 31, 1999, January 1, 2000–December 31, 2002, January 1, 2003–December 31, 2006 and January 1, 2007–December 31, 2010. The four intervals were chosen so as to yield approximately equal sample sizes across the intervals and so that the last interval would consist of studies published a sufficiently long time after publication of the CONSORT statement extension in March 2004 to allow time to incorporate the guidelines into planned statistical analyses as well as time for analysis and publication. 95% confidence intervals were obtained as Clopper-Pearson binomial exact confidence intervals. Fisher’s exact test was used to test for differences in proportions between time intervals.

To explore whether failure to use appropriate methods was associated with a higher probability of a finding of statistical significance, we classified each article as having one or more statistically significant results for cancer screening outcomes, and compared these proportions between the “appropriate” and “not appropriate” method groups using Fisher’s

exact test. One study was omitted from this analysis because the authors considered their results ambiguous.

All analyses were conducted using Stata 11 [18].

Results

Table 1 summarizes the articles reporting outcome analyses of cluster randomized trials of cancer screening interventions that we identified through our search. We identified 50 articles meeting our inclusion criteria, including 14 articles reporting the outcome of a breast screening intervention trial, 6 reporting a cervical cancer screening intervention trial, 20 reporting colorectal cancer screening interventions, 4 reporting joint interventions for both breast and cervical cancer screening and 6 reporting interventions for screening all three cancer sites.

In 25 of the studies, the randomization units were health care providers or facilities, most commonly medical practices (n=10) or physicians (n=9). There were 21 community-based trials, in which the units of randomization were communities or neighborhoods (n=8), housing units (n=3) or sites where community members congregate such as churches, clubs or senior centers (n=10). Two studies randomized by worksite, and two by family.

Review of the methodology used for analyzing the cancer screening outcome in each study indicated that 72% (36/50, exact 95% confidence interval 58%–84%) of the studies used methods appropriate for cluster randomized trials. Among high influence (n=22) versus low influence (n=28) articles, 77% (55%–92%) versus 68% (48%–84%) used appropriate methods; the difference was not significant ($p=.54$, Fisher's exact test).

Figure 1 presents smoothed estimates of the running proportions of articles reporting appropriate methods over the study period of 1995 to 2010, for all articles (Figure 1(a)), high influence articles (Figure 1(b)), and low influence articles (Figure 1(c)). All three figures suggest a rising trend followed by a plateau then a decline. Logistic regression confirmed a significant quadratic trend in the overall sample ($p=.03$) and yielded results consistent with quadratic trends in both the high and low influence articles (both $p=.08$). For articles in high influence journals, the proportion appeared to reach a higher plateau than for articles in low influence journals. During this plateau, which ran from 2000 to 2007, all articles in high influence journals used appropriate methods. The smoothed estimates suggested that the proportion for low influence articles fell to a lower level by 2010 than that for high influence articles.

Table 2 provides the proportions of studies that used appropriate analytic methods by publication time interval and results of tests for differences in proportions over time. The proportions in the intervals 1995–1999, 2000–2002, 2003–2006 and 2007–2010 were 0.55, 0.82, 0.92 and 0.60, respectively. No differences in proportions were significant at the 0.05 level; however, the results were consistent with a peak in the 2003–2006 time frame ($p=.06$) and drop from 2003–2006 to 2007–2010 ($p=.08$).

Of the articles using appropriate methods, the proportion reporting a statistically significant finding 0.72 (95% exact confidence interval 0.53–0.86). For studies not using appropriate methods, the proportion was 0.76 (0.50–0.93). The difference between the two groups was not statistically significant ($p = .99$ by Fisher's exact test).

Studies not using appropriate analytic methods most often erred by treating participant outcomes as independent rather than correlated. Of the 36 studies using analytic methods that accounted for clustering, the most commonly used method was mixed models (31%,

11/35), followed by generalized estimating equations (26%, 9/33). Four studies stated that they used a robust variance estimator, four used clusters as the unit of analysis and three used survey data methods. Other methods included the Cochran-Mantel-Haenszel chi-square test, adjusted z-test [19], and permutation tests. In several cases, the papers included statements to the effect that the standard errors were adjusted for clustering or a hierarchical model was used to account for the design, but the particular method used was not stated explicitly.

Only seven studies reported ICC values for the primary outcome [20–26]; one of these studies reported the ICC for the control group only [21]. An additional two studies [27, 28] reported the design effect, which is equal to $1 + \text{ICC} \times (\text{average cluster size} - 1)$. One study [29] reported ICCs for baseline data but not outcome data. Table 3 provides a compilation of ICC values for outcome data from the trials. The percentages of studies reporting ICCs for outcomes by time interval were 9% (1/11) in 1995–1999, 18% (2/11) in 2000–2002, 38% (5/13) in 2003–2006 and 7% (1/15) in 2007–2010. Fisher's exact test gave a p-value of .07 for a difference between the last two time intervals. While these numbers are small, the pattern of proportions is consistent with the time trend for use of appropriate methods.

Discussion

Our investigation suggests that many cluster randomized trials of cancer screening interventions have had deficiencies in the application of correct statistical procedures for the outcome analysis, and that a trend towards increased adoption of appropriate methods in the late 1990s and early 2000s may not have continued after the mid-2000s. In particular, we did not find evidence to support an increase in the use of appropriate methods or reporting of the ICC after publication of the CONSORT statement extension to cluster randomized trials in March 2004. A particularly striking finding is that only a single study in the 2007–2010 time interval reported an ICC. Overall, differences in the rates of use of appropriate methods between articles in high influence versus low influence journals were not remarkable. However, there was a period in the mid-2000s when all articles in high influence journals used appropriate methods.

It is instructive to examine our findings in the context of earlier reviews of the methodological quality of cluster randomized trials. Past reviews of cluster randomized trial methodology for health-related studies are summarized in Table 4. The studies that examined time trends had findings consistent with increased adoption of appropriate methods over periods ranging to the mid-2000s, which is consistent with our findings. We speculate that ongoing methodological reviews highlighting deficiencies and increased dissemination of proper methods in seminal books, e.g., [2, 7] were responsible for the increasing trend through the mid-2000's. None of these reviews included articles beyond 2008 and thus these studies would not have captured the downward trend in recent years that was suggested by our review. Our review adds to the literature because it includes more recent publications through 2010, many of them published well after the 2004 CONSORT extension for cluster randomized trials.

Our findings have several implications. First, since failure to account for clustering generally inflates the Type I error rate, it is likely that some non-equivalence trials of cancer screening interventions have overstated the statistical significance of the intervention effect. In particular, there may be instances in which an intervention was declared effective when a careful analysis adjusting for clustering would have failed to find an effect. Thus the evidence base for some cancer screening interventions may be overstated. In our review, the rates of any finding of statistically significance for cancer screening outcomes were similar between studies using appropriate methods and those that did not; however, our measure

may have been insufficiently sensitive to detect a difference. We found that it was difficult to extract from the articles important information that would have permitted a more sensitive measure to be used (e.g., exact number of tests conducted, p-values for each test).

Our findings also have implications for the design of future trials of cancer screening interventions. Determination of the sample size needed for a planned cluster randomized trial to achieve the desired level of power generally requires an estimate of the ICC (see, e.g., [1, 2, 30]). An underestimated ICC will result in an underpowered study, and an overestimated ICC will cause a study to be overly large and wasteful of resources. Published literature is an important source of plausible ICC values for the planning of future studies [3, 6]. Omission of ICC values from published studies represents a lost opportunity to help in the efficient and effective design of future studies. A recent Journal of the National Cancer Institute monograph providing ICC estimates for cancer screening outcomes highlights the importance of published ICCs as a resource for other investigators who are planning studies [31]. Given this paucity of reported ICCs in the trial literature, investigators may wish to consider using such methods as sample size adaptive designs [32] as a way of dealing with uncertainty in the magnitude of the ICC that will be realized in a trial.

Three articles reported ICCs or design effects separately for each arm of the trial [26–28], and the arm-specific ICCs were quite different. A reanalysis of the trial reported by Mishra et al. [23] also found differences in the ICCs by study condition, with an ICC of 0.06 in the control condition and 0.34 in the intervention condition [33]. Thus in addition to accounting for correlation among participants, investigators planning or analyzing data from cluster randomized trials may also wish to consider the possibility of differences in the magnitude of the ICC between study conditions.

We found that it is difficult to systematically identify cluster randomized trials in the published literature. There is no standard nomenclature for describing a cluster randomized design; searches on terms such as “cluster randomized” or “group randomized” miss the many studies that do not use these terms anywhere in their abstracts, titles, keywords or text; this lead to our decision to broadly capture articles reporting randomized trials then scrutinize each such article to determine whether the trial was cluster randomized. Identification of a trial as cluster randomized often entailed a careful reading of the methods section of a paper, and even then, in some cases the randomization scheme was still unclear. A recent article by Taljaard et al [34] describes the challenges of identifying reports of cluster randomized trials in the literature in detail. Conformance with the CONSORT statement extension, which calls for identifying a trial as cluster randomized in the title and abstract, would remedy this problem.

The categorization of a trial as cluster randomized can be ambiguous. We excluded two studies that were time-randomized [35, 36] and thus might have involved some clustering effects attributable to time period. Some studies involved assigning individuals to small groups and randomizing these groups (e.g., [37, 38]). While it is debatable whether such studies should be termed “cluster randomized,” such designs do involve correlation among participants within the same group, a feature that should be accounted for in study design and analysis. Since we were not able to identify such studies systematically and because the necessity of accounting for correlation among participants in such trials has not been widely discussed or disseminated, we did not include such studies in our review.

We restricted our literature search to studies that had receipt of breast, cervical or colorectal cancer screening (yes/no) as an outcome. Thus our pool of studies did not include cancer screening interventions that did not have receipt of screening as an outcome, although cluster randomization is also used in many such interventions (e.g., [39]).

Conclusions

In conclusion, our work shows that there is a need for renewed attention to the methodological quality of cluster randomized trials. While we confined our investigation to breast, cervix and colorectal cancer screening intervention trials, it is likely that other intervention and clinical studies share similar methodological and reporting deficiencies. Future reports of cancer screening intervention trials should adhere to the CONSORT statement and its extension to cluster randomized trials and include clear descriptions of the randomization scheme, unit of analysis and the statistical methods used to account for correlated data, and report the ICC to allow for better planning of future studies.

Abbreviations

BR	breast
CE	cervical
CR	colorectal
CONSORT	Consolidated Standards of Reporting Trials
FOBT	fecal occult blood test
ICC	intraclass correlation coefficient

Acknowledgments

Financial Support

This research was support by National Institutes of Health grants CA137827, CA16042, and CA109091.

Role of the Funding Source

The funding source had no involvement in the conduct of the research.

References

1. Hayes, RJ.; Moulton, LH. Cluster Randomised Trials. Boca Raton, Florida: CRC Press; 2009.
2. Donner, A.; Klar, N. Design and Analysis of Cluster Randomization Trials in Health Research New York. New York: Oxford University Press; 2000.
3. Klar N, Donner A. Current and future challenges in the design and analysis of cluster randomization trials. *Stat Med.* 2001; 20:3729–3740. [PubMed: 11782029]
4. Campbell MJ, Donner A, Klar N. Developments in cluster randomized trials and Statistics in Medicine. *Stat Med.* 2007; 26:2–19. [PubMed: 17136746]
5. Campbell MK, Mollison J, Grimshaw JM. Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. *Stat Med.* 2001; 20:391–399. [PubMed: 11180309]
6. 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. [PubMed: 14998806]
7. Murray, DM. Design and Analysis of Group-Randomized Trials. New York, New York: Oxford University Press; 1998.
8. 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. [PubMed: 2084005]
9. Divine GW, Brown JT, Frazier LM. The unit of analysis error in studies about physicians' patient care behavior. *J Gen Intern Med.* 1992; 7:623–629. [PubMed: 1453246]

10. Simpson JM, Klar N, Donnor A. Accounting for cluster randomization: a review of primary prevention trials, 1990 through 1993. *Am J Public Health*. 1995; 85:1378–1383. [PubMed: 7573621]
11. Chuang JH, Hripesak G, Jenders RA. Considering clustering: a methodological review of clinical decision support system studies. *Proc AMIA Symp*. 2000:146–150. [PubMed: 11079862]
12. Isaakidis P, Ioannidis JP. Evaluation of cluster randomized controlled trials in sub-Saharan Africa. *Am J Epidemiol*. 2003; 158:921–926. [PubMed: 14585770]
13. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. *BMJ*. 2004; 328:702–708. [PubMed: 15031246]
14. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996; 276:637–639. [PubMed: 8773637]
15. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001; 357:1191–1194. [PubMed: 11323066]
16. Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol*. 1978; 108:100–102. [PubMed: 707470]
17. Sasieni P, Royston P, Cox NJ. Symmetric nearest neighbor linear smoothers. *Stata Journal*. 2005; 5:285.
18. StataCorp. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP; 2009.
19. 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. discussion 300-1. [PubMed: 8030631]
20. Slater JS, Ha CN, Malone ME, McGovern P, Madigan SD, Finnegan JR, et al. A randomized community trial to increase mammography utilization among low-income women living in public housing. *Prev Med*. 1998; 27:862–870. [PubMed: 9922069]
21. Richards SH, Bankhead C, Peters TJ, Austoker J, Hobbs FD, Brown J, et al. Cluster randomised controlled trial comparing the effectiveness and cost-effectiveness of two primary care interventions aimed at improving attendance for breast screening. *J Med Screen*. 2001; 8:91–98. [PubMed: 11480450]
22. Segura J. A Randomized Controlled Trial Comparing Three Invitation Strategies in a Breast Cancer Screening Program. *Prev Med*. 2001; 33:325–332. [PubMed: 11570837]
23. Mishra SI, Bastani R, Crespi CM, Chang LC, Luce PH, Baquet CR. Results of a Randomized Trial to Increase Mammogram Usage among Samoan Women. *Cancer Epidemiol Biomarkers Prev*. 2007; 16:2594–2604. [PubMed: 18086763]
24. Campbell MK, James A, Hudson MA, Carr C, Jackson E, Oates V, et al. Improving multiple behaviors for colorectal cancer prevention among African American church members. *Health Psychology*. 2004; 23:492–502. [PubMed: 15367069]
25. Federici A, Rossi PG, Borgia P, Bartolozzi F, Farchi S, Guasticchi G. The immunochemical faecal occult blood test leads to higher compliance than the guaiac for colorectal cancer screening programmes: a cluster randomized controlled trial. *J Med Screen*. 2005; 12:83–88. [PubMed: 15949119]
26. Holloway RM, Wilkinson C, Peters TJ, Russell I, Cohen D, Hale J, et al. Cluster-randomised trial of risk communication to enhance informed uptake of cervical screening. *Brit J Gen Pract*. 2003; 53:620–625. [PubMed: 14601338]
27. Federici A, Marinacci C, Mangia M, Borgia P, Giorgirossi P, Guasticchi G. Is the type of test used for mass colorectal cancer screening a determinant of compliance? A cluster-randomized controlled trial comparing fecal occult blood testing with flexible sigmoidoscopy. *Cancer Detect Prev*. 2006; 30:347–353. [PubMed: 16965874]
28. Ferreira MR. Health Care Provider-Directed Intervention to Increase Colorectal Cancer Screening Among Veterans: Results of a Randomized Controlled Trial. *J Clin Oncol*. 2005; 23:1548–1554. [PubMed: 15735130]
29. Jensen H, Svanholm H, Stovring H, Bro F. A primary healthcare-based intervention to improve a Danish cervical cancer screening programme: a cluster randomised controlled trial. *J Epidemiol Community Health*. 2009; 63:510–515. [PubMed: 19228681]

30. Campbell MK, Thomson S, Ramsay CR, MacLennan GS, Grimshaw JM. Sample size calculator for cluster randomized trials. *Comput Biol Med.* 2004; 34:113–125. [PubMed: 14972631]
31. Hade EM, Murray DM, Pennell ML, Rhoda D, Paskett ED, Champion VL, et al. Intraclass correlation estimates for cancer screening outcomes: estimates and applications in the design of group-randomized cancer screening studies. *J Natl Cancer Inst Monogr.* 2010; 2010:97–103. [PubMed: 20386058]
32. Chow, S-C.; Chang, M. *Adaptive Design Methods in Clinical Trials.* Boca Raton, FL: Chapman & Hall/CRC; 2007.
33. Crespi CM, Wong WK, Mishra SI. Using second-order generalized estimating equations to model heterogeneous intraclass correlation in cluster-randomized trials. *Stat Med.* 2009; 28:814–827. [PubMed: 19109804]
34. Taljaard M, McGowan J, Grimshaw JM, Brehaut JC, McRae A, Eccles MP, et al. Electronic search strategies to identify reports of cluster randomized trials in MEDLINE: low precision will improve with adherence to reporting standards. *BMC Med Res Methodol.* 2010; 10
35. Potter MB, Phengrasamy L, Hudes ES, McPhee SJ, Walsh JME. Offering Annual Fecal Occult Blood Tests at Annual Flu Shot Clinics Increases Colorectal Cancer Screening Rates. *Ann Fam Med.* 2009; 7:17–23. [PubMed: 19139445]
36. Ward JE, Proude EM. Evaluation of doctors' reminders in emergency departments to encourage cervical screening. *Australian New Zealand J Public Health.* 1999; 23:95–98. [PubMed: 10083698]
37. Maxwell A, Bastani R, Vida P, Warda US. Results of a randomized trial to increase breast and cervical cancer screening among Filipino American women. *Prev Med.* 2003; 37:102–109. [PubMed: 12855209]
38. Maxwell AE, Bastani R, Danao LL, Antonio C, Garcia GM, Crespi CM. Results of a Community-Based Randomized Trial to Increase Colorectal Cancer Screening Among Filipino Americans. *Am J Public Health.* 2010; 100:2228–2234. [PubMed: 20864724]
39. Zheng S, Chen K, Liu X, Ma X, Yu H, Yao K, et al. Cluster randomization trial of sequence mass screening for colorectal cancer. *Dis Colon Rectum.* 2003; 46:51–58. [PubMed: 12544522]
40. Urban N, Taplin SH, Taylor VM, Peacock S, Anderson G, Conrad D, et al. Community organization to promote breast cancer screening among women ages 50–75. *Prev Med.* 1995; 24:477–484. [PubMed: 8524722]
41. Clover K, Redman S, Forbes J, SansonFisher R, Callaghan T. Two sequential randomized trials of community participation to recruit women for mammographic screening. *Prev Med.* 1996; 25:126–134. [PubMed: 8860277]
42. Dietrich AJ, Tobin JN, Sox CH, Cassels AN, Negron F, Younge RG, et al. Cancer early-detection services in community health centers for the underserved - A randomized controlled trial. *Arch Fam Med.* 1998; 7:320–327. [PubMed: 9682685]
43. Hodge FS, Stubbs HA, Gurgin V, Fredericks L. Cervical cancer screening - Knowledge, attitudes, and behavior of American Indian women. *Cancer.* 1998; 83:1799–1804.
44. Kinsinger LS, Harris R, Qaqish B, Strecher V, Kaluzny A. Using an office system intervention to increase breast cancer screening. *J Gen Intern Med.* 1998; 13:507–514. [PubMed: 9734786]
45. Manfredi C, Czaja R, Freels S, Trubitt M, Warnecke R, Lacey L. Prescribe for health. Improving cancer screening in physician practices serving low-income and minority populations. *Arch Fam Med.* 1998; 7:329–337. [PubMed: 9682686]
46. Bastani R, Maxwell AE, Bradford C, Das IP, Yan KX. Tailored risk notification for women with a family history of breast cancer. *Prev Med.* 1999; 29:355–364. [PubMed: 10564627]
47. Bonevski B, Sanson-Fisher RW, Campbell E, Carruthers A, Reid ALA, Ireland M. Randomized controlled trial of a computer strategy to increase general practitioner preventive care. *Prev Med.* 1999; 29:478–486. [PubMed: 10600428]
48. Powe BD, Weinrich S. An intervention to decrease cancer fatalism among rural elders. *Oncol Nurs Forum.* 1999; 26:583–588. [PubMed: 10214599]
49. Tilley BC, Vernon SW, Myers R, Glanz K, Lu M, Hirst K, et al. The Next Step Trial: impact of a worksite colorectal cancer screening promotion program. *Prev Med.* 1999; 28:276–283. [PubMed: 10072746]

50. Andersen MR, Yasui Y, Meischke H, Kuniyuki A, Etzioni R, Urban N. The effectiveness of mammography promotion by volunteers in rural communities. *Am J Prev Med.* 2000; 18:199–207. [PubMed: 10722985]
51. Thompson N. A Randomized Controlled Trial of a Clinic-Based Support Staff Intervention to Increase the Rate of Fecal Occult Blood Test Ordering. *Prev Med.* 2000; 30:244–251. [PubMed: 10684748]
52. Allen JD, Stoddard AM, 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. [PubMed: 11291370]
53. Hancock L. Effect of a Community Action Intervention on Cervical Cancer Screening Rates in Rural Australian Towns: The CART Project. *Prev Med.* 2001; 32:109–117. [PubMed: 11162336]
54. Powe BD. Promoting fecal occult blood testing in rural African American women. *Cancer Pract.* 2002; 10:139–146. [PubMed: 11972568]
55. Reuben DB, Bassett LW, Hirsch SH, Jackson CA, Bastani R. A randomized clinical trial to assess the benefit of offering on-site mobile mammography in addition to health education for older women. *AJR Am J Roentgenol.* 2002; 179:1509–1514. [PubMed: 12438046]
56. Taylor VM, Jackson JC, Yasui Y, Kuniyuki A, Acorda E, Marchand A, et al. Evaluation of an outreach intervention to promote cervical cancer screening among Cambodian American women. *Cancer Detect Prev.* 2002; 26:320–327. [PubMed: 12430637]
57. Vinker S, Nakar S, Rosenberg E, Kitai E. The role of family physicians in increasing annual fecal occult blood test screening coverage: a prospective intervention study. *Isr Med Assoc J.* 2002; 4:424–425. [PubMed: 12073414]
58. Zhu K. An Intervention Study on Screening for Breast Cancer among Single African-American Women Aged 65 and Older. *Prev Med.* 2002; 34:536–545. [PubMed: 11969355]
59. Luckmann R. A randomized trial of telephone counseling to promote screening mammography in two HMOs. *Cancer Detect Prev.* 2003; 27:442–450. [PubMed: 14642552]
60. Roetzheim RG. A Randomized Controlled Trial to Increase Cancer Screening Among Attendees of Community Health Centers. *The Annals of Family Medicine.* 2004; 2:294–300.
61. Ruffin M. Interventions fail to increase cancer screening rates in community-based primary care practices. *Prev Med.* 2004; 39:435–440. [PubMed: 15313081]
62. Braun K, Fong M, Kaanoi M, Kamaka M, Gotay C. Testing a culturally appropriate, theory-based intervention to improve colorectal cancer screening among native Hawaiians. *Prev Med.* 2005; 40:619–627. [PubMed: 15850857]
63. Hughes K, Leggett B, Del Mar C, Croese J, Fairley S, Masson J, et al. Guaiac versus immunochemical tests: faecal occult blood test screening for colorectal cancer in a rural community. *Austral New Zealand J Public Health.* 2005; 29:358–364.
64. Roetzheim RG. Long-term Results From a Randomized Controlled Trial to Increase Cancer Screening Among Attendees of Community Health Centers. *Ann Fam Med.* 2005; 3:109–114. [PubMed: 15798035]
65. Walsh JME, Salazar R, Terdiman JP, Gildengorin G, Perez-Stable EJ. Promoting Use of Colorectal Cancer Screening Tests. Can We Change Physician Behavior? *J Gen Intern Med.* 2005; 20:1097–1101. [PubMed: 16423097]
66. Thompson B, Coronado G, Chen L, Islas I. Celebremos La Salud! A Community Randomized Trial of Cancer Prevention (United States). *Cancer Causes Control.* 2006; 17:733–746. [PubMed: 16633921]
67. Dinshaw K, Mishra G, Shastri S, Badwe R, Kerkar R, Ramani S, et al. Determinants of Compliance in a Cluster Randomised Controlled Trial on Screening of Breast and Cervix Cancer in Mumbai, India. *Oncology.* 2007; 73:145–153. [PubMed: 18408401]
68. Stephens JH, Moore JWE. Can targeted intervention in CRC patients? relatives influence screening behaviour? A pilot study. *Colorectal Disease.* 2007 0:070621084454019-???
69. Hiatt RA, Pasick RJ, Stewart S, Bloom J, Davis P, Gardiner P, et al. Cancer Screening for Underserved Women: The Breast and Cervical Cancer Intervention Study. *Cancer Epidemiol Biomarkers Prev.* 2008; 17:1945–1949. [PubMed: 18708383]

70. Jandorf L, Bursac Z, Pulley L, Trevino M, Castillo A, Erwin DO. Breast and cervical cancer screening among Latinas attending culturally specific educational programs. *Prog Community Health Partnersh.* 2008; 2:195–204. [PubMed: 20208198]
71. Lane DS, Messina CR, Cavanagh MF, Chen JJ. A provider intervention to improve colorectal cancer screening in county health centers. *Med Care.* 2008; 46:S109–S116. [PubMed: 18725822]
72. Ling BS, Schoen RE, Trauth JM, Wahed AS, Eury T, Simak DM, et al. Physicians encouraging colorectal screening: a randomized controlled trial of enhanced office and patient management on compliance with colorectal cancer screening. *Arch Intern Med.* 2009; 169:47–55. [PubMed: 19139323]
73. Ma GX, Shive S, Tan Y, Gao WZ, Rhee J, Park M, et al. Community-based colorectal cancer intervention in underserved Korean Americans. *Cancer Epidemiol.* 2009; 33:381–386. [PubMed: 19914880]
74. Sequist TD, Zaslavsky AM, Marshall R, Fletcher RH, Ayanian JZ. Patient and physician reminders to promote colorectal cancer screening: a randomized controlled trial. *Arch Intern Med.* 2009; 169:364–371. [PubMed: 19237720]
75. Atlas SJ, Grant RW, Lester WT, Ashburner JM, Chang Y, Barry MJ, et al. A Cluster-Randomized Trial of a Primary Care Informatics-Based System for Breast Cancer Screening. *J Gen Intern Med.* 2010
76. Blumenthal DS, Smith SA, Majett CD, Alema-Mensah E. A trial of 3 interventions to promote colorectal cancer screening in African Americans. *Cancer.* 2010; 116:922–929. [PubMed: 20052732]
77. Liu CY, Xia HO, Isaman DM, Deng W, Oakley D. Nursing clinical trial of breast self-examination education in China. *Int Nurs Rev.* 2010; 57:128–134. [PubMed: 20487485]
78. Morgan PD, Fogel J, Tyler ID, Jones JR. Culturally Targeted Educational Intervention to Increase Colorectal Health Awareness among African Americans. *Journal of Health Care for the Poor and Underserved.* 2010; 21:132–147. [PubMed: 20675951]
79. Ornstein S, Nemeth LS, Jenkins RG, Nietert PJ. Colorectal cancer screening in primary care: translating research into practice. *Med Care.* 2010; 48:900–906. [PubMed: 20808257]
80. MacLennan GS, Ramsay CR, Mollison J, Campbell MK, Grimshaw JG, Thomas RE. Room for improvement in the reporting of cluster randomised trials in behaviour change research. *Control Clin Trials.* 2003; 24:69s–70s.
81. Varnell SP, Murray DM, Janega JB, Blitstein JL. Design and analysis of group-randomized trials: A review of recent practices. *American Journal of Public Health.* 2004; 94:393–399. [PubMed: 14998802]
82. Murray DM, Pals SL, Blitstein JL, Alfano CM, Lehman J. Design and analysis of group-randomized trials in cancer: A review of current practices. *J Natl Cancer Inst.* 2008; 100:483–491. [PubMed: 18364501]
83. Bowater RJ, Abdelmalik SME, Lilford RJ. The methodological quality of cluster randomised controlled trials for managing tropical parasitic disease: a review of trials published from 1998 to 2007. *Trans Roy Soc Trop Med Hyg.* 2009; 103:429–436. [PubMed: 19232658]
84. Handlos LN, Chakraborty H, Sen PK. Evaluation of cluster-randomized trials on maternal and child health research in developing countries. *Trop Med Int Health.* 2009; 14:947–956. [PubMed: 19563429]

Figure 1. Smoothed running proportions of articles reporting appropriate methods, 1995 to 2010
Symmetric nearest neighbor smoothing was used to obtain smoothed estimates of the running proportion of studies using appropriate methods. The binary indicator for use of appropriate methods was logit transformed and a span of 0.4 was used. Values of 0 and 1 are jittered and plotted below and above the smoothed curves, respectively.

Table 1

Studies reporting outcome analyses of cluster randomized trials of breast, cervical or colorectal cancer screening interventions

First author	Year	Randomization units	Cancer site	Reference
Urban	1995	Communities	BR	[40]
Clover	1996	Communities	BR	[41]
Dietrich	1998	Health centers	BR,CV,CR	[42]
Hodge	1998	Clinics	CV	[43]
Kinsinger	1998	Medical practices	BR	[44]
Manfredi	1998	Medical practices	BR,CV,CR	[45]
Slater	1998	Apartment buildings	BR	[20]
Bastani	1999	Families	BR	[46]
Bonevski	1999	Physicians	CV	[47]
Powe	1999	Senior centers	CR	[48]
Tilley	1999	Worksites	CR	[49]
Andersen	2000	Communities	BR	[50]
Thompson	2000	Clinical care teams	CR	[51]
Allen	2001	Worksites	BR,CV	[52]
Hancock	2001	Communities	CV	[53]
Richards	2001	Medical practices	BR	[21]
Segura	2001	Apartment buildings	BR	[22]
Powe	2002	Senior centers	CR	[54]
Reuben	2002	Various community-based sites	BR	[55]
Taylor	2002	Neighborhoods	CV	[56]
Vinker	2002	Physicians	CR	[57]
Zhu	2002	Housing complexes	BR	[58]
Holloway	2003	Medical practices	CV	[26]
Luckmann	2003	Medical practices	BR	[59]
Campbell	2004	Churches	CR	[24]
Roetzheim	2004	Clinics	BR,CV,CR	[60]
Ruffin	2004	Medical practices	BR,CV,CR	[61]
Braun	2005	Civic clubs	CR	[62]
Federici	2005	Physicians	CR	[25]
Ferreira	2005	Clinics	CR	[28]
Hughes	2005	Medical practices	CR	[63]
Roetzheim	2005	Clinics	BR,CV,CR	[64]
Walsh	2005	Physicians	CR	[65]
Federici	2006	Medical practices	CR	[27]
Thompson	2006	Communities	BR,CV,CR	[66]
Dinshaw	2007	Communities	BR,CV	[67]
Mishra	2007	Churches	BR	[23]
Stephens	2007	Families	CR	[68]

First author	Year	Randomization units	Cancer site	Reference
Hiatt	2008	Clinics	BR,CV	[69]
Jandorf	2008	Various community-based sites	BR,CV	[70]
Lane	2008	Health centers	CR	[71]
Jensen	2009	Physicians	CV	[29]
Ling	2009	Physicians	CR	[72]
Ma	2009	Churches	CR	[73]
Sequist	2009	Physicians	CR	[74]
Atlas	2010	Medical practices	BR	[75]
Blumenthal	2010	Various community-based sites	CR	[76]
Liu	2010	Communities	BR	[77]
Morgan	2010	Churches, community-based organizations	CR	[78]
Ornstein	2010	Medical practices	CR	[79]

BR, breast; CR, colorectal; CV, cervix

Table 2

Proportions of articles using appropriate analytic methods for cancer screening outcomes by publication time interval

Publication time interval	Proportion of studies using appropriate methods for cancer screening outcome	Exact 95% confidence interval	P-value, difference of proportions compared to previous interval	P-value, difference of proportions compared to 1995–1999
By interval:				
1995–1999	0.55 (6/11)	0.23–0.83		
2000–2002	0.82 (9/11)	0.48–0.98	.36	.36
2003–2006	0.92 (12/13)	0.64–0.99	.58	.06
2007–2010	0.60 (9/15)	0.32–0.84	.08	.99
Overall:				
1995–2010	0.72 (36/50)	0.58–0.84		

P-values for differences of proportions were obtained using Fisher's exact test.

Table 3

Intraclass correlation coefficients reported in cancer screening intervention trials

Reference	Year	Cluster type	Outcome	ICC
Slater [20]	1998	Apartment buildings	Self-reported receipt of mammogram	−0.0015
Segura [22]	2001	Apartment buildings	Attendance for mammogram	0.065
Campbell [24]	2004	Churches	Fecal occult blood test screening	<0.01
Mishra [23]	2007	Churches	Self-reported receipt of mammogram	0.19
Richards [21]	2001	Medical practices	Attendance for breast screening	0.023 in control group
Holloway [26]	2003	Medical practices	Receipt of Pap smear	0.25 in intervention group, 0.0018 in control group
Federici [27]	2006	Medical practices	Receipt of FOBT or flexible sigmoidoscopy	0.008 in FOBT arm, <0.001 in flexible sigmoidoscopy arm ^a
Federici [25]	2005	Physicians	Receipt of FOBT	0.056
Ferreira [28]	2005	Physicians	Receipt of colorectal cancer screening tests	0.077 in intervention group, 0.033 in control group ^a

FOBT, fecal occult blood test; ICC, intraclass correlation coefficient

^a Design effect provided in original article; ICC was computed as (design effect − 1)/(average cluster size − 1)

Table 4

Summary of previous published reviews of methodology used in health-related cluster randomized trials

Reference	Type of studies reviewed	Publication years included	Percent of studies accounting for clustering in analysis	Time trends (when reported)
Donner [8]	Non-therapeutic intervention trials	1979–1989	50% (8/16)	1979–1985: 17% (1/6) 1986–1989: 70% (7/10)
Divine [9]	Physician patient-care behavior studies	1980–1990	30% (16/54)	No time trend
Simpson [10]	Primary prevention trials published in <i>American Journal of Public Health</i> and <i>Preventive Medicine</i>	1990–1993	57% (12/21)	NA
Chuang [11]	Clinical decision support system studies	1975–1998	58% (14/24)	Odds ratio 1.12 (95% CI 0.97–1.30) for one-year increase in year of publication
Isaakidis [12]	Sub-Saharan Africa intervention trials	1973–2001	37% (19/51)	1973–1995: 24% (6/25) 1996–2001: 50% (13/26)
MacLennan [80]	Guideline implementation studies	1966–2000	58% (87/149)	Odds ratio 1.08 (p=.025) for one-year increase in year of publication
Varnell [81]	Trials published in <i>American Journal of Public Health</i> and <i>Preventive Medicine</i>	1998–2002	54% (32/59)	NA
Murray [82]	Cancer prevention and control trials	2002–2006	45% (34/75)	NA
Bowater [83]	Tropical parasitic disease intervention trials	1998–2007	43% (15/35)	1998–2002: 33% (6/18) 2003–2007: 53% (9/17)
Handlos [84]	Trials of maternal and child health care interventions in developing countries	1998–2008	80% (28/35)	1998–2002: 38% (3/8) 2003–2008: 67% (18/27)