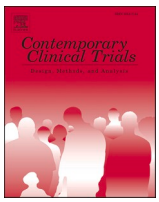
Contemporary Clinical Trials 123 (2022) 106974

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial



Cluster-randomized trials of cancer screening interventions: Has use of appropriate statistical methods increased over time?

Catherine M. Crespi *, Kevin Ziehl

Department 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, United States

ARTICLE INFO

Keywords:
Cluster randomized trials
Intraclass correlation coefficient
CONSORT
Cancer screening intervention trials

ABSTRACT

Background: In a cluster randomized trial, groups of individuals (e.g., clinics, schools) are randomized to conditions. The design and analysis of cluster randomized trials can require more care than individually randomized trials. Past reviews have noted deficiencies in the use of appropriate statistical methods for such trials.
Methods: We reviewed cluster randomized trials of cancer screening interventions published 1995–2019 to determine whether appropriate statistical methods had been used for sample size calculation and outcome analysis and whether they reported intraclass correlation coefficient (ICC) values. This work expanded a previous review of articles published 1995–2010.
Results: Our search identified 88 articles published 1995–2020 that reported outcomes of cluster randomized trials of breast, cervix, and colorectal cancer screening interventions. There was increased reporting of the trials' sample size calculations over time, with the percentage increasing from 31% in 1995–2004 to 77% in 2014–2019. However, the percentage of calculations failing to account for cluster randomization did not change over time and was 17% of studies in 2014–2019. There was a nonsignificant trend towards increased use of outcome analysis methods that accounted for the cluster randomized design. However, in lower impact journals, use of appropriate analysis methods was only 80% in 2014–2019. Only 33% of studies reported ICC values in 2014–2019.
Conclusion: For cluster randomized trials with cancer screening outcomes, there have been improvements in the reporting of sample size calculations but methodological and reporting deficiencies persist. Efforts to disseminate, adopt and report the use of appropriate statistical methodologies are still needed.

1. Introduction

Cluster randomized trials (CRT) are trials that randomize groups or "clusters" of individuals to different treatment conditions, as opposed to randomizing individuals. CRTs are commonly used to evaluate non-drug interventions, such as service delivery interventions, group-based behavioral interventions or cancer screening programs. The clusters can be, for example, clinics, hospitals, health care providers, churches or entire communities.

A cluster randomized design may be selected for several reasons [1–3]. Cluster randomization can help prevent contamination of the control condition with the intervention, since all cluster members receive the same condition. The intervention may be designed for implementation at the group level, e.g., group therapy or a clinic-wide change in procedures. In some cases, a cluster randomized design may be chosen because it is less costly or logistically easier to implement the intervention at the cluster level.

The design and analysis of a CRT requires attention to the clustered nature of the data. Members of the same cluster generally share physical, geographic or other characteristics, which makes members of the same cluster more similar to one another as compared to members of different clusters. As a result, in cluster randomized trials, the outcomes of individuals within the same cluster are not independent but rather are correlated with each other, a feature of the data that must be accounted for in the design and analysis of the trial. The degree of within-cluster correlation of the outcome is measured by the intracluster correlation

Abbreviations: BR, breast; CE, cervical; CR, colorectal; CRT, cluster randomized trial; CONSORT, Consolidated Standards of Reporting Trials; FOBT, fecal occult blood test; ICC, intraclass correlation coefficient.

* Corresponding author.
E-mail address: ccrespi@ucla.edu (C.M. Crespi).

https://doi.org/10.1016/j.cct.2022.106974
Received 14 July 2022; Received in revised form 30 September 2022; Accepted 20 October 2022
Available online 4 November 2022

coefficient (ICC), which can be interpreted as the Pearson correlation of two observations within the same cluster [1,4]. A positive ICC leads to an increase in the variance of estimator of the intervention effect [1,5]. Neglecting this additional variance can lead to artificially small p-values and potentially misleading conclusions [6,1]. The complexity of the analysis of a CRT has not always been appreciated, as indicated by widespread deficiencies in the conduct and reporting CRTs in the past [7–12].

Previous work by our group reviewed cluster randomized trials of breast, cervix and colorectal cancer screening interventions published in 1995–2010 to determine whether the use of appropriate statistical methods for the outcome analysis had increased over time [13]. Results were suggestive of a rise in the use of appropriate methods from 1995–1999 (55% of articles) to 2003–2006 (92%), followed by a decline in 2007–2010 (60%).

In this paper, we update the previous review to include CRTs of cancer screening interventions published through 2019. Our objective was to characterize recent trends in the use of proper statistical methods in these CRTs. In addition to investigating whether such studies reported the use of outcome analysis methods appropriate for a cluster-randomized design, we expanded the investigation to assess whether the studies reported a sample size calculation and whether the calculation used methods appropriate for cluster-randomized designs. We further investigated whether the use of appropriate methods was associated with journal impact and whether studies reported ICC values. The reporting of ICC values is useful for investigators planning similar studies, and is recommended by the CONSORT guidelines for CRTs [14].

2. Methods

2.1. Literature search

For an article to be included in the current review, it had to meet the following criteria: (1) breast, cervical or colon cancer screening as a primary outcome; (2) published between January 1, 2011 and December 31, 2019; (3) written in English; (4) used parallel groups cluster randomized trial design; and (5) article was primary report of the trial outcomes. This is a slight modification of the criteria used in the previous review [13], which did not restrict to primary outcomes or primary reports of trials. Since our goal was to include all articles meeting the criteria, including papers that did not self-identify as a CRT and as a result may have used inappropriate methods, our strategy was to first identify all articles reporting a randomized trial with a cancer screening outcome and then review each article to determine whether the design was a parallel groups cluster randomized trial meeting our criteria. We conducted literature searches in PubMed using the same search strategy as was done for the previous review; the search term combination was

((("clinical trial"[Publication Type])AND "english'ı[Language])) AND
("2011/01/01'ı[Date − MeSH] : "3000'ı[Date − MeSH]))) AND
(randomized trial SITE cancer screening[Text Word])

where SITE was specified as "colorectal", "cervical" or "breast". An end date of December 31, 2019 was used. Supplementary searches were conducted by identifying common MeSH terms in already identified publications and using them to expand the search base. We further attempted to find outcome papers for publications that only presented preliminary results or the study protocol.

2.2. Variables

All identified articles were reviewed to determine whether they met the inclusion criteria. Articles determined to meet the criteria were reviewed by two statisticians (the authors) in order to extract: whether a sample size or power calculation was included and whether it considered the CRT design; the statistical method used to analyze the cancer screening outcome; and whether an ICC value was reported for the cancer screening outcome.

Appropriateness of sample size calculation method. Cluster randomization has an impact on statistical power by inflating the variance of the estimated intervention effect. This variance inflation factor, which is also called the design effect, can be calculated as $1 + (m-1)ICC$, where m is the average cluster size [1]. Variance inflation can also be quantified using the coefficient of variation of the cluster-level outcomes [3], which is the standard deviation of the cluster-level outcomes divided by the mean cluster-level outcome. An article was classified as using an appropriate sample size calculation method if it reported the use of an intraclass correlation coefficient, variance inflation factor, design effect or coefficient of variation as part of the sample size calculation. Vague reports such as mentioning that the calculation "considered the clustered design" were also classified as "appropriate". Articles were categorized as (1) appropriate sample size calculation reported; (2) inappropriate sample size calculation reported; and (3) no sample size calculation reported.

Appropriateness of outcome analysis method. In order to classify the statistical methods used in the outcome analysis as appropriate or not, we used Murray et al. [12] as a reference for common CRT analysis techniques and also consulted authoritative texts [1,3,15] and references cited in the articles that we reviewed. Appropriate analytic methods can be classified as either cluster-level methods or individual-level methods. Cluster-level methods use outcomes at the level of the cluster as the units of analysis. Summary statistics are computed at the cluster level (for example, proportion of cluster members with the outcome) and then the data set of summary statistics is analyzed using methods for independent units. The analysis applied to the summary statistics could be a two-sample t test, Wilcoxon rank sum test or permutation test. Individual-level analysis methods use the outcomes of individuals as the units of analysis, and include the following methods:

Mixed models. The statistical model includes random effects that induce correlation of outcomes within cluster. For a binary outcome such as receipt of cancer screening, a mixed effects logistic regression model is typically used. References for this method include [3,15,5].

Generalized estimating equations (GEE). This method involves estimating an outcome model (e.g., a logistic regression model) in conjunction with a working covariance structure that estimates the within-cluster correlation and is used to adjust the standard errors for clustering. References for this method include [1,3].

Cluster robust standard error estimation. Standard errors are computed using a formula that accounts for correlation of observations within cluster; see [16,17].

Sampling design weighting. The observations are regarded as being cluster-sampled. Survey sampling methods are used to derive sample weights that are used in the analysis to adjust for clustering. References for this method include [18].

Adjusted chi-square test. A correction for clustering is applied to the chi-square statistic for independent observations. This approach is designed for binary outcomes; see [1].

Articles that stated that their method accounted for the clustered design were classified as appropriate even if no additional information was provided to conclusively determine the appropriateness of their approach.

Journal impact. We determined the Source Normalized Impact per Paper (SNIP) for each journal in which the articles were published. The SNIP is similar to the traditional Journal Impact Factor, which is the average number of times that articles from the journal published in the past two years have been cited. However, the SNIP weights citations based on the total number of citations in a subject field. Thus the impact of a single citation is given higher weight in subject areas where citations are less frequent, and vice versa. This corrects for differences in citation practices between scientific fields. The SNIP is also based on three years of cited publications rather than two. Further details can be found at https://lib.guides.umd.edu/bibliometrics. SNIP metrics were accessed

through the Centre for Science and Technology Studies [19].

We dichotomized the SNIP values at the median value; journals with SNIP value above the median were classified as "higher impact" and those below were classified as "lower impact". For several older journals, SNIP values were not available; however, other journal impact metrics indicated that these journals could be appropriately classified as lower impact.

2.3. Statistical analysis

Articles from the previous review were combined with articles from the updated search to create the dataset used to conduct analyses. To examine time trends, the full study period was broken into three shorter time periods with approximately equal numbers of articles (1995–2004, 2005–2013, 2014–2019). Proportions of articles with four outcomes of interest – sample size calculation reported, appropriate sample size calculation (among those reporting a calculation), appropriate outcome analysis method, and ICC reported (among those with appropriate analysis method) – were calculated for each time period. We tested for linear trends by fitting a logistic regression model with the outcome as the dependent variable and a linear term in calendar year of publication as the explanatory variable. All analyses were repeated after stratifying on lower versus higher journal impact. Differences between lower and higher journal impact for the 2014–2019 time period were assessed using Fisher exact tests. Results were deemed statistically significant if a two-sided *p*-value was below.05. Analyses were conducted in Stata/SE 17.0.

3. Results

Our search initially identified 90 articles related to breast cancer screening, 299 articles related to colorectal cancer screening, and 139 related to cervical cancer screening published between January 1, 2011 and December 31, 2019. After review, 40 were found to meet the inclusion criteria. Re-review of the previous set of articles published 1995–2010 revealed that two articles [20,21] did not meet the more stringent criteria used in this review and they were dropped from the analysis. The total number of articles included in this analysis was 88.

Characteristics of the 40 studies from 2011–2019 are summarized in Table 1. Of these articles, 25 reported a colorectal cancer screening outcome, 9 reported a cervical cancer screening outcome, 3 reported a breast cancer screening outcome, one had both a breast cancer and cervical cancer screening outcome, and two reported on all three screening outcomes. There were 25 different types of randomization units, with the most common being physicians (n = 6), followed by clinics, health workers, practices, and geographical regions. In one study, the participants were sent CRC screening invitations over a specified study time period and randomization was by week of invitation [22].

The proportions of studies with any sample size calculation, an appropriate sample size calculation, an appropriate outcome analysis method, and ICC reporting for three successive time periods, 1995–2004, 2005–2014 and 2015–2019, are reported in Table 2. The proportion of articles reporting a sample size calculation increased from 0.31 in the earliest period to 0.77 in the most recent period ($p = .002$, linear trend test). Among articles reporting a sample size calculation, the proportion reporting one that accounted for clustering remained similar across the study period, with 88% in 1995–2004 and 2005–2014 and 83% in 2014–2019 ($p = .840$ for linear trend). The use of appropriate outcome analysis methods increased from 0.73 to 0.90, but the linear trend was not statistically significant ($p = .091$). Among articles using appropriate analysis methods, ICC reporting remained low throughout the study period, ranging from 0.21 in the earlier period to 0.33 in the most recent period, with no detectable trend ($p = .403$).

In total, 80% (70/88) of articles reported using an outcome analysis method for accounted for the cluster randomized design. The remainder

Table 1
Studies published 2011–2019 reporting outcome analyses of cluster randomized trials of breast, cervical or colorectal cancer screening interventions.

First author	Year	Randomization units	Cancer site	Reference
Acera	2017	Health Centers	CV	[24]
Arrossi	2015	Health Workers	CV	[25]
Atlas	2011	Practices	BR	[26]
Atlas	2014	Primary Care Practices	BR, CV, CR	[27]
Aubin-Auger	2016	Practices	CR	[28]
Barthe	2015	Physicians	CR	[29]
Birkenfeld	2011	Primary Care Clinics	CR	[30]
Clouston	2014	Physicians	CR	[31]
Cuaresma	2018	Health Educators	CR	[32]
De Mil	2018	Geographical Regions	CR	[33]
Decker	2013	Geographical Regions	CV	[34]
Dignan	2014	Primary Care Practices	CR	[35]
Dodd	2019	Day of Invitation	CR	[36]
Fang	2017	Churches	CV	[37]
Fernandez	2015	Neighborhoods	CR	[38]
Ghaffari	2019	Health Centers	BR	[39]
Guillaume	2017	Geographical Regions	CR	[40]
Guiriguet	2016	Physicians	CR	[41]
Han	2017	Churches	BR, CV	[42]
Huchko	2018	Communities	CV	[43]
Jo	2017	Health Workers	CR	[44]
Kitchener	2018	Medical Practices	CV	[45]
Krok-Schoen	2015	Clinics	CR	[46]
Le Breton	2016	Physicians	CR	[47]
Leone	2016	Churches	CR	[48]
Lo	2014	Invitation week	CR	[22]
Ma	2015	Communities	CV	[49]
Maxwell	2016	Organizations	CR	[50]
Nguyen	2017	Health Workers	CR	[51]
Nguyen	2015	Lay Health Workers	CR	[52]
Price-Haywood	2014	Clinics	BR, CV, CR	[53]
Rat	2017	General Practitioners	CR	[54]
Sadler	2011	Salons	BR	[55]
Shaw	2013	Practices	CR	[56]
Sun	2018	Primary Care Physicians	CR	[57]
Tinmouth	2015	Physicians	CR	[58]
Tong	2017	Lay Health Educators	CR	[59]
Wang	2018	Physicians	CR	[60]
Wong	2019	Community Centers	CV	[61]
Zehbe	2016	Communities	CV	[62]

Abbreviations: BR, breast cancer; CR, colorectal cancer; CV, cervical cancer.

Table 2
Proportions of trials reporting sample size calculation, appropriate sample size calculation, appropriate outcome analysis method, and ICC reporting over time.

	1995–2004	2005–2014	2015–2019	*p*, linear trend
Sample size calculation				
Reported	0.31 (8/26)	0.50 (16/32)	0.77 (23/30)	.002
Appropriate	0.88 (7/8)	0.88 (14/16)	0.83 (19/23)	.840
Outcome analysis method				
Appropriate	0.73 (19/26)	0.75 (24/32)	0.90 (27/30)	.091
Intraclass correlation				
Reported	0.21 (4/19)	0.29 (7/24)	0.33 (9/27)	.403

P-values for linear trend were obtained from logistic regression models regressing the outcome on calendar year as a linear term. Analysis of whether sample size calculation was appropriate was restricted to articles reporting a sample size calculation. Analysis of intraclass correlation reporting was restricted to articles that reported using appropriate analysis methods.

all reported using methods for independent observations. Among articles using appropriate methods, the most commonly used method to account for clustering was mixed models (36%, 25/70), followed by generalized estimating equations (26%, 18/70), cluster-level analysis (13%, 9/70) and cluster robust standard errors (11%, 8/70). Other methods included sample design weighting ($n = 4$) and the adjusted chi-square test ($n = 3$). For three articles, it was stated that the analysis accounting for clustering but the method used was not clear.

The SNIP scores of the journals in which the articles were published ranged from 0.04 to 11.13, with a median of 1.27. Results stratified by lower versus higher journal impact (SNIP below or above median) are presented in Table 3. Among articles in lower impact journals, the proportion reporting a sample size calculation increased from 0.40 to 0.67 ($p = .141$, linear trend test). Among those reporting one, the percentage with an appropriate calculation varied from 83% to 100% to 90% across the time periods, with no trend detected ($p = .649$). The overall proportion reporting use of appropriate outcome analysis methods was 0.70 (31/44), with no significant time trend ($p = .434$). Among those with appropriate analysis methods, the overall proportion reporting an ICC was 0.29, also with no time trend ($p = .873$).

Among articles in higher impact journals, the proportion reporting a sample size calculation increased significantly from 0.18 in the earliest period to 0.87 in the most recent period ($p = .003$, linear trend test). Among those reporting a calculation, the overall proportion with calculations adjusted for clustering was 80% (20/25), with no detectable time trend ($p = .596$). The proportions using appropriate methods for outcome analysis rose from 0.82 in 1995–2004 to 1.00 in 2015–2019 ($p = .098$ for linear trend). Reporting of ICCs increased from 0.11 to 0.33 ($p = .288$ for linear trend).

With regard to difference in linear time trends between articles in lower and higher impact journals, tests for interactions between lower/higher impact and calendar year in logistic regression models did not detect any significant differences in any of the four outcomes (all $p > .20$). Fisher exact tests comparing articles in lower versus higher impact journals in the most recent time period found no statistically significant differences in the four outcomes.

Table 3
Stratification by journal impact: Proportions of trials reporting sample size calculation, appropriate sample size calculation, appropriate outcome analysis method, and ICC reporting over time.

	1995–2004	2005–2014	2015–2019	*p*, linear trend
Lower impact				
Sample size calculation				
Reported	0.40 (6/15)	0.43 (6/14)	0.67 (10/15)	.141
Appropriate	0.83 (5/6)	1.00 (6/6)	0.90 (9/10)	.649
Outcome analysis method				
Appropriate	0.67 (10/15)	0.64 (9/14)	0.80 (12/15)	.434
Intraclass correlation				
Reported	0.30 (3/10)	0.22 (2/9)	0.33 (4/12)	.873
Higher impact				
Sample size calculation				
Reported	0.18 (2/11)	0.56 (10/18)	0.87 (13/15)	.003
Appropriate	1.00 (2/2)	0.80 (8/10)	0.78 (10/13)	.596
Outcome analysis method				
Appropriate	0.82 (9/11)	0.83 (15/18)	1.00 (15/15)	.098
Intraclass correlation				
Reported	0.11 (1/9)	0.33 (5/15)	0.33 (5/15)	.288

Lower and higher journal impact was defined by dichotomizing the journal Source Normalized Impact per Paper (SNIP) score. *P*-values for linear trend were obtained from logistic regression models regressing the outcome on calendar year as a linear term. Analysis of whether sample size calculation was appropriate was restricted to articles reporting a sample size calculation. Analysis of intraclass correlation reporting was restricted to articles that reported using appropriate analysis methods.

4. Discussion

Prior studies have noted deficiencies in the use and reporting of appropriate statistical methods for the design and analysis of cluster randomized trials. In this review of cluster randomized trials of cancer screening interventions published 1995–2019, we found that there had been improvement over time in some areas but a lack of improvement in others.

It is important to account for the cluster randomized design when calculating the sample size required to achieve the desired level of power for a planned cluster randomized trial; otherwise, a study may be underpowered and inconclusive. Transparent reporting of the sample size calculation when reporting the results of a study is considered a benchmark of quality and allows for critical appraisal of the trial and potential for biased results [23]. In this study, we found a significant increase over time in the proportion of studies that reported a sample size calculation, with the proportion more than doubling from 0.31 to 0.77. This increase was driven by a dramatic increase among higher impact journals, together with a lesser increase among lower impact journals. This suggests that study authors, peer reviewers and/or journal editors are paying more attention to the importance of reporting this information. However, in 2014–2019, 23% of studies still did not report a sample size calculation. Some of this lack of reporting may be attributable to a lack of awareness of best practice reporting guidelines such as the CONSORT statement [14]. However, even if authors are aware of reporting guidelines, the word and page count limits of many journals may force authors to prioritize the reporting of some information over other information.

While the reporting of sample size calculations increased over time, there was no time trend in whether those calculations accounted for the cluster randomized design or not. Overall, during 1995–2019, about 15% of articles that reported a sample size calculation did not indicate that the cluster randomized design was accounted for. Furthermore, articles in higher impact journals did not use appropriate sample size methods at a higher rate than those in lower impact journals. This suggests that there may be an enduring proportion of researchers who are not well-schooled in the design of cluster randomized trials, and that better outreach and training are needed.

In cluster randomized trials, analyzing the outcome data using a statistical technique that accounts for the cluster randomized design is critical for obtaining accurate inference. Failure to account for clustering is likely to deflate the standard errors and thereby overstate the statistical significance of findings. Our examination of the use of appropriate methods for outcome analysis in these cluster randomized trials found that the time trends were consistent with increased use of appropriate analysis methods over time, although the increase was not statistically significant. In the latest time period examined, 2014–2019, the use of appropriate analysis methods was 90% overall. When stratified by journal impact, the percentages were 80% for articles in lower impact journals and 100% in higher impact journals. This suggests that there is a greater potential for the statistical significance of findings to be overstated in cluster randomized trials report in the lower impact journals. We also found that some descriptions of the statistical analyses used were vague and did not clearly specify the statistical technique that was used. This highlights a need for statistical expertise when writing, reviewing and editing papers reporting clinical trials.

The proportion of articles that reported ICC values increased about

50% from 1995–2004 to 2014–2019, rising from 0.21 to 0.33. However, even in recent years, only a minority of articles reported ICC values. Furthermore, articles in higher impact journals were not more likely to report ICC values than those in lower impact journals. There could be several explanations for the lack of reporting. Many software commands used to analyze clustered data, such as those fitting mixed logistic models and generalized estimating equations, do not include ICC values in their output. Rather, extra programming steps are needed to compute ICC values, and statistical expertise is needed to do so. Lack of awareness of best practice reporting guidelines and constraints on page and word counts could also hinder ICC reporting. Reliable estimates of intraclass correlation coefficients are critical for planning future studies. Thus this lack of reporting could hinder accurate design of future CRTs. The CONSORT guidelines for CRTs recommend reporting ICCs [6]. Greater adherence to these guidelines would benefit cancer prevention and other research areas.

We learned from this review that it is not always easy to identify cluster randomized trials in the literature. About 90% (80/88) of the articles that were included in this trial could be identified as using a cluster randomized design based on information in the title and/or abstract. However, many of these articles did not use the terms "cluster randomized" or "group randomized", and the use of cluster randomization had to be inferred from other information that was provided, e.g., "worksites were randomized to conditions" or "providers randomized to the intervention…" combined with a patient-level outcome analysis. About 10% (8/88) of the included articles could not be identified as cluster randomized trials based on the title or abstract and could only be determined to be CRTs based on a full reading of the paper. Overall, we had to thoroughly review over 500 articles in order to positively identify 40 of them as reporting cluster randomized trials meeting our inclusion criteria. The CONSORT 2010 statement extension to cluster randomised trials recommends that a CRT be identified as cluster randomized in the title [14]. Increased and clear reporting of trial design information, especially in the title, are needed to facilitate literature searches and the finding of relevant studies.

This work has limitations. Our results may not be generalizable to all cluster randomized trials since our effort was limited to those with cancer screening outcomes. Additionally, due to the lack of standardization of descriptive terms utilized in the literature and inherent difficulty of identifying cluster randomized trials that did not self-identify as such, we might have missed relevant articles. This may have created a bias towards overestimating the use and reporting of appropriate statistical methods.

5. Conclusions

This work builds on an earlier assessment to examine trends over a 25-year period. Our results indicate that despite continuing experience in the field with cluster randomized trials, many articles still do not use appropriate statistical methods and/or do not report clear information on their design and analysis methods. More outreach and training are needed to remedy these deficiencies and ensure the robustness and accuracy of the scientific literature. It is important that researchers receive, when appropriate, more specialized training in the design of clinical trials and are made aware of alternative study designs and what statistical implications their choices might have, as well as the importance of thorough reporting on study design and analysis methods.

Funding sources

This work was supported by National Institutes of Health/National Cancer Institute grant P30 CA016042.

Data availability

Data will be made available on request.

References

[1] A. Donner, N. Klar, Design and Analysis of Cluster Randomization Trials in Health Research, Arnold, London, 2000.
[2] N. Klar, A. Donner, Current and future challenges in the design and analysis of cluster randomization trials, Stat. Med. 20 (24) (2001) 3729–3740.
[3] R.J. Hayes, L.H. Moulton, Cluster Randomised Trials, CRC Press, 2017.
[4] S.M. Eldridge, O.C. Ukoumunne, J.B. Carlin, The Intra-Cluster Correlation Coefficient in Cluster Randomized Trials: A Review of Definitions, Int. Stat. Rev. 77 (2009) 378–394.
[5] C.M. Crespi, Cluster randomized trials, in: S. Halabi, S. Michiels (Eds.), Textbook of Clinical Trials in Oncology: A Statistical Perspective, CRC Press/Taylor and Francis Group, 2019.
[6] M.K. Campbell, J. Mollison, J.M. Grimshaw, Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size, Stat. Med. 20 (3) (2001) 391–399.
[7] A. Donner, K.S. Brown, P. Brasher, A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979–1989, Int. J. Epidemiol. 19 (4) (1990) 795–800.
[8] G.W. Divine, J.T. Brown, L.M. Frazier, The unit of analysis error in studies about physicians' patient care behavior, J. Gen. Intern. Med. 7 (6) (1992) 623–629.
[9] J.M. Simpson, N. Klar, A. Donnor, Accounting for cluster randomization: a review of primary prevention trials, 1990 through 1993, Am. J. Public Health 85 (10) (1995) 1378–1383.
[10] J.-H. Chuang, G. Hripcsak, R.A. Jenders, Considering clustering: a methodological review of clinical decision support system studies, in: Proceedings of the AMIA Symposium, American Medical Informatics Association, 2000, p. 146.
[11] P. Isaakidis, J.P.A. Ioannidis, Evaluation of cluster randomized controlled trials in sub-Saharan Africa, Am. J. Epidemiol. 158 (9) (2003) 921–926.
[12] D.M. Murray, S.L. Pals, S.M. George, A. Kuzmichev, G.Y. Lai, J.A. Lee, R.L. Myles, S.M. Nelson, Design and analysis of group-randomized trials in cancer: a review of current practices, Prev. Med. 111 (2018) 241–247.
[13] C.M. Crespi, A.E. Maxwell, S. Wu, Cluster randomized trials of cancer screening interventions: are appropriate statistical methods being used? Contemp. Clin. Trials 32 (4) (2011) 477–484.
[14] M.K. Campbell, G. Piaggio, D.R. Elbourne, D.G. Altman, Consort 2010 statement: extension to cluster randomised trials, BMJ 345 (2012), e5661.
[15] M. Moerbeek, S. Teerenstra, Power Analysis of Trials with Multilevel Data, CRC Press/Taylor & Francis Group, 2016.
[16] R.H. Brook, J.E. Ware, W.H. Rogers, E.B. Keeler, A.R. Davies, C.A. Donald, G. A. Goldberg, K.N. Lohr, P.C. Masthay, J.P. Newhouse, Does free care improve adults' health? Results from a randomized controlled trial, N. Engl. J. Med. 309 (23) (1983) 1426–1434.
[17] M.J. Campbell, A. Donner, N. Klar, Developments in cluster randomized trials and Statistics in Medicine, Stat. Med. 26 (1) (2007) 2–19.
[18] E.S. Lee, R.N. Forthofer, Analyzing Complex Survey Data, Sage Publications, 2005.
[19] Centre for Science and Technology Studies.https://www.journalindicators.com/. Last accessed 10 December 2020.
[20] M.K. Campbell, A. James, M.A. Hudson, C. Carr, E. Jackson, V. Oates, S. Demissie, D. Farrell, I. Tessaro, Improving multiple behaviors for colorectal cancer prevention among African American church members, Health Psychol. 23 (5) (2004) 492–502.
[21] R.G. Roetzheim, L.K. Christman, P.B. Jacobsen, A.B. Cantor, J. Schroeder, R. Abdulla, S. Hunter, T.N. Chirikos, J.P. Krischer, A randomized controlled trial to increase cancer screening among attendees of community health centers, Ann. Fam. Med. 2 (4) (2004) 294–300.
[22] S.H. Lo, A. Good, P. Sheeran, G. Baio, S. Rainbow, G. Vart, C. von Wagner, J. Wardle, Preformulated implementation intentions to promote colorectal cancer screening: a cluster-randomized trial, Health Psychol. 33 (9) (2014) 998.
[23] D.G. Altman, K.F. Schulz, D. Moher, M. Egger, F. Davidoff, D. Elbourne, P. C. Gøtzsche, T. Lang, the CONSORT Group, The revised CONSORT statement for reporting randomized trials: explanation and elaboration, Ann. Intern. Med. 134 (8) (2001) 663–694.
[24] A. Acera, J.M. Manresa, D. Rodriguez, A. Rodriguez, J.M. Bonet, M. Trapero-Bertran, P. Hidalgo, N. Sanchez, S. De Sanjose, Increasing cervical cancer screening coverage: a randomised, community-based clinical trial, PLOS One 12 (1) (2017), e0170371.
[25] S. Arrossi, L. Thouyaret, R. Herrero, A. Campanera, A. Magdaleno, M. Cuberli, P. Barletta, R. Laudi, L. Orellana, EMA Study Team, Effect of self-collection of HPV DNA offered by community health workers at home visits on uptake of screening for cervical cancer (the EMA study): a population-based cluster-randomised trial, Lancet Glob. Health 3 (2) (2015) e85–e94.
[26] S.J. Atlas, R.W. Grant, W.T. Lester, J.M. Ashburner, Y. Chang, M.J. Barry, H. C. Chueh, A cluster-randomized trial of a primary care informatics-based system for breast cancer screening, J. Gen. Intern. Med. 26 (2) (2011) 154–161.
[27] S.J. Atlas, A.H. Zai, J.M. Ashburner, Y. Chang, S. Percac-Lima, D.E. Levy, H. C. Chueh, R.W. Grant, Non-visit-based cancer screening using a novel population management system, J. Am. Board Fam. Med. 27 (4) (2014) 474–485.
[28] I. Aubin-Auger, C. Laouénan, J. Le Bel, A. Mercier, D. Baruch, J.P. Lebeau, A. Youssefian, T. Le Trung, L. Peremans, P. Van Royen, Efficacy of communication skills training on colorectal cancer screening by GPs: a cluster randomised controlled trial, Eur. J. Cancer Care 25 (1) (2016) 18–26.
[29] J. Barthe, E. Perrodeau, S. Gilberg, P. Ravaud, C. Ghasarossian, F. Marchand-Buttin, J. Deyra, H. Falcoff, Impact of a doctor's invitation on participation in colorectal cancer screening: a cluster randomized trial, Am. J. Med. 128 (9) (2015) 1024–e1.

[30] S. Birkenfeld, R.G. Belfer, M. Chared, A. Vilkin, M. Barchana, I. Lifshitz, D. Fruchter, D. Aronski, R. Balicer, Y. Niv, et al., Factors affecting compliance in faecal occult blood testing: a cluster randomized study of the faecal immunochemical test versus the guaiac faecal occult test, J. Med. Screen. 18 (3) (2011) 135–141.
[31] K. Clouston, A. Katz, P.J. Martens, J. Sisler, D. Turner, M. Lobchuk, S. McClement, G. Crow, CIHR/CCMB Team in Primary Care Oncology (PCO-NET), et al., Does access to a colorectal cancer screening website and/or a nurse-managed telephone help line provided to patients by their family physician increase fecal occult blood test uptake?: Results from a pragmatic cluster randomized controlled trial, BMC Cancer 14 (1) (2014) 263.
[32] C.F. Cuaresma, A.U. Sy, T.T. Nguyen, R.C.S. Ho, G.L. Gildengorin, J.Y. Tsoh, A. M. Jo, E.K. Tong, M. Kagawa-Singer, S.L. Stewart, Results of a lay health education intervention to increase colorectal cancer screening among Filipino Americans: a cluster randomized controlled trial, Cancer 124 (2018) 1535–1542.
[33] R. De Mil, E. Guillaume, L. Guittet, O. Dejardin, V. Bouvier, C. Pornet, V. Christophe, A. Notari, H. Delattre-Massy, C. De Seze, et al., Cost-effectiveness analysis of a navigation program for colorectal cancer screening to reduce social health inequalities: a French cluster randomized controlled trial, Value Health 21 (6) (2018) 685–691.
[34] K.M. Decker, D. Turner, A.A. Demers, P.J. Martens, P. Lambert, D. Chateau, Evaluating the effectiveness of cervical cancer screening invitation letters, J. Women's Health 22 (8) (2013) 687–693.
[35] M. Dignan, B. Shelton, S.A. Slone, C. Tolle, S. Mohammad, N. Schoenberg, K. Pearce, E. Van Meter, G. Ely, Effectiveness of a primary care practice intervention for increasing colorectal cancer screening in Appalachian Kentucky, Prev. Med. 58 (2014) 70–74.
[36] N. Dodd, M. Carey, E. Mansfield, C. Oldmeadow, T.-J. Evans, Testing the effectiveness of a general practice intervention to improve uptake of colorectal cancer screening: a randomised controlled trial, Aust. N. Z. J. Public Health 43 (5) (2019) 464–469.
[37] C.Y. Fang, G.X. Ma, E.A. Handorf, Z. Feng, Y. Tan, J. Rhee, S.M. Miller, C. Kim, H. S. Koh, Addressing multilevel barriers to cervical cancer screening in Korean American women: a randomized trial of a community-based intervention, Cancer 123 (6) (2017) 1018–1026.
[38] M.E. Fernández, L.S. Savas, C.C. Carmack, W. Chan, D.R. Lairson, T.L. Byrd, K. M. Wilson, S.R. Arvey, S. Krasny, S.W. Vernon, A randomized controlled trial of two interventions to increase colorectal cancer screening among Hispanics on the Texas-Mexico border, Cancer Causes Control 26 (1) (2015) 1–10.
[39] M. Ghaffari, S.N. Esfahani, S. Rakhshanderou, P.H. Koukamari, Evaluation of health belief model-based intervention on breast cancer screening behaviors among health volunteers, J. Cancer Educ. 34 (5) (2019) 904–912.
[40] E. Guillaume, O. Dejardin, V. Bouvier, R. De Mil, C. Berchi, C. Pornet, V. Christophe, A. Notari, H.D. Massy, C. De Seze, et al., Patient navigation to reduce social inequalities in colorectal cancer screening participation: a cluster randomized controlled trial, Prev. Med. 103 (2017) 76–83.
[41] C. Guiriguet, L. Muñoz-Ortiz, A. Burón, I. Rivero, J. Grau, C. Vela-Vallespín, M. Vilarrubí, M. Torres, C. Hernández, L. Méndez-Boo, et al., Alerts in electronic medical records to promote a colorectal cancer screening programme: a cluster randomised controlled trial in primary care, Br. J. Gen. Pract. 66 (648) (2016) e483–e490.
[42] H.-R. Han, Y. Song, M. Kim, H.K. Hedlin, K. Kim, H. Ben Lee, D. Roter, Breast and cervical cancer screening literacy among Korean American women: a community health worker–led intervention, Am. J. Public Health 107 (1) (2017) 159–165.
[43] M.J. Huchko, S. Ibrahim, C. Blat, C.R. Cohen, J.S. Smith, R.A. Hiatt, E. Bukusi, Cervical cancer screening through human papillomavirus testing in community health campaigns versus health facilities in rural western Kenya, Int. J. Gynecol. Obstet. 141 (1) (2018) 63–69.
[44] A.M. Jo, T.T. Nguyen, S. Stewart, M.J. Sung, G. Gildengorin, J.Y. Tsoh, E.K. Tong, P. Lo, C. Cuaresma, A. Sy, et al., Lay health educators and print materials for the promotion of colorectal cancer screening among Korean Americans: a randomized comparative effectiveness study, Cancer 123 (14) (2017) 2705–2715.
[45] H. Kitchener, M. Gittins, M. Cruickshank, C. Moseley, S. Fletcher, R. Albrow, A. Gray, L. Brabin, D. Torgerson, E.J. Crosbie, et al., A cluster randomized trial of strategies to increase uptake amongst young women invited for their first cervical screen: the STRATEGIC trial, J. Med. Screen. 25 (2) (2018) 88–98.
[46] J.L. Krok-Schoen, G.S. Young, M.L. Pennell, P.L. Reiter, M.L. Katz, D.M. Post, C. M. Tatum, E.D. Paskett, Testing interventions to motivate and educate (TIME): a multi-level intervention to improve colorectal cancer screening, Prev. Med. Rep. 2 (2015) 306–313.
[47] J. Le Breton, É. Ferrat, C. Attali, S. Bercier, P. Le Corvoisier, Z. Brixi, K. Veerabudun, V. Renard, S. Bastuji-Garin, Effect of reminders mailed to general practitioners on colorectal cancer screening adherence: a cluster-randomized trial, Eur. J. Cancer Prev. 25 (5) (2016) 380–387.
[48] L.A. Leone, M. Allicock, M.P. Pignone, J.F. Walsh, L.-S. Johnson, J. Armstrong-Brown, C.C. Carr, A. Langford, A. Ni, K. Resnicow, et al., Cluster randomized trial of a church-based peer counselor and tailored newsletter intervention to promote colorectal cancer screening and physical activity among older African Americans, Health Educ. Behav. 43 (5) (2016) 568–576.
[49] G.X. Ma, C. Fang, Y. Tan, Z. Feng, S. Ge, C. Nguyen, Increasing cervical cancer screening among Vietnamese Americans: a community-based intervention trial, J. Health Care Poor Underserved 26 (2 Suppl) (2015) 36.
[50] A.E. Maxwell, L.L. Danao, R.T. Cayetano, C.M. Crespi, R. Bastani, Implementation of an evidence-based intervention to promote colorectal cancer screening in community organizations: a cluster randomized trial, Transl. Behav. Med. 6 (2) (2016) 295–305.
[51] T.T. Nguyen, J.Y. Tsoh, K. Woo, S.L. Stewart, G.M. Le, A. Burke, G. Gildengorin, R. J. Pasick, J. Wang, E. Chan, et al., Colorectal cancer screening and Chinese Americans: efficacy of lay health worker outreach and print materials, Am. J. Prev. Med. 52 (3) (2017) e67–e76.
[52] B.H. Nguyen, S.L. Stewart, T.T. Nguyen, N. Bui-Tong, S.J. McPhee, Effectiveness of lay health worker outreach in reducing disparities in colorectal cancer screening in Vietnamese Americans, Am. J. Public Health 105 (10) (2015) 2083–2089.
[53] E.G. Price-Haywood, J. Harden-Barrios, L.A. Cooper, Comparative effectiveness of audit-feedback versus additional physician communication training to improve cancer screening for patients with limited health literacy, J. Gen. Intern. Med. 29 (8) (2014) 1113–1121.
[54] C. Rat, C. Pogu, D. Le Donné, C. Latour, G. Bianco, F. Nanin, A. Cowppli-Bony, A. Gaultier, J.-M. Nguyen, Effect of physician notification regarding nonadherence to colorectal cancer screening on patient participation in fecal immunochemical test cancer screening: a randomized clinical trial, JAMA 318 (9) (2017) 816–824.
[55] G.R. Sadler, C.M. Ko, P. Wu, J. Alisangco, S.F. Castañeda, C. Kelly, A cluster randomized controlled trial to increase breast cancer screening among African American women: the black cosmetologists promoting health program, J. Natl Med. Assoc. 103 (8) (2011) 735–745.
[56] E.K. Shaw, P.A. Ohman-Strickland, A. Piasecki, S.V. Hudson, J.M. Ferrante, R. R. McDaniel, P.A. Nutting, B.F. Crabtree, Effects of facilitated team meetings and learning collaboratives on colorectal cancer screening rates in primary care practices: a cluster randomized trial, Ann. Fam. Med. 11 (3) (2013) 220–228.
[57] A. Sun, J.Y. Tsoh, E.K. Tong, J. Cheng, E.A. Chow, S.L. Stewart, T.T. Nguyen, A physician-initiated intervention to increase colorectal cancer screening in Chinese patients, Cancer 124 (2018) 1568–1575.
[58] J. Tinmouth, J. Patel, P.C. Austin, N.N. Baxter, M.C. Brouwers, C. Earle, C. Levitt, Y. Lu, M. Mackinnon, L. Paszat, et al., Increasing participation in colorectal cancer screening: Results from a cluster randomized trial of directly mailed FOBT kits to previous nonresponders, Int. J. Cancer 136 (6) (2015) E697–E703.
[59] E.K. Tong, T.T. Nguyen, P. Lo, S.L. Stewart, G.L. Gildengorin, J.Y. Tsoh, A.M. Jo, M. L. Kagawa-Singer, A.U. Sy, C. Cuaresma, et al., Lay health educators increase colorectal cancer screening among Hmong Americans: a cluster randomized controlled trial, Cancer 123 (1) (2017) 98–106.
[60] H.-Y. Wang, G.X. Ma, W. Liang, Y. Tan, K.H. Makambi, R. Dong, S.W. Vernon, S.-P. Tu, J.S. Mandelblatt, et al., Physician intervention and Chinese Americans' colorectal cancer screening, Am. J. Health Behav. 42 (1) (2018) 13–26.
[61] C.L. Wong, K.C. Choi, B.M.H. Law, D.N.S. Chan, W.K.W. So, Effects of a community health worker-led multimedia intervention on the uptake of cervical cancer screening among South Asian women: a pilot randomized controlled trial, Int. J. Environ. Res. Public Health 16 (17) (2019) 3072.
[62] I. Zehbe, R. Jackson, B. Wood, B. Weaver, N. Escott, A. Severini, M. Krajden, L. Bishop, K. Morrisseau, G. Ogilvie, et al., Community-randomised controlled trial embedded in the Anishinaabek cervical cancer screening study: human papillomavirus self-sampling versus Papanicolaou cytology, BMJ Open 6 (10) (2016).