Preventive Medicine 111 (2018) 241–247

Contents lists available at ScienceDirect

Preventive Medicine

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

Review Article

Design and analysis of group-randomized trials in cancer: A review of current practices☆

David M. Murray[a,*], Sherri L. Pals[b,1], Stephanie M. George[a], Andrey Kuzmichev[c], Gabriel Y. Lai[d], Jocelyn A. Lee[e], Ranell L. Myles[a], Shakira M. Nelson[f]

[a] Office of Disease Prevention, Division of Program Coordination Planning and Strategic Initiatives, Office of the Director, National Institutes of Health, Bethesda, MD, United States
[b] Health Informatics, Data Management, and Statistics Branch, Division of Global HIV and Tuberculosis, Center for Global Health, US Centers for Disease Control and Prevention, Atlanta, GA, United States
[c] Office of the Surgeon General, Office of the Assistant Secretary for Health, Department of Health and Human Services, United States
[d] Environmental Epidemiology Branch, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Rockville, MD, United States
[e] Project Genomics Evidence Neoplasia Information Exchange (GENIE), Executive Office, American Association for Cancer Research, Philadelphia, PA, United States
[f] Scientific Programs, American Association for Cancer Research, Philadelphia, PA, United States

A B S T R A C T

The purpose of this paper is to summarize current practices for the design and analysis of group-randomized trials involving cancer-related risk factors or outcomes and to offer recommendations to improve future trials.

We searched for group-randomized trials involving cancer-related risk factors or outcomes that were published or online in peer-reviewed journals in 2011–15. During 2016–17, in Bethesda MD, we reviewed 123 articles from 76 journals to characterize their design and their methods for sample size estimation and data analysis.

Only 66 (53.7%) of the articles reported appropriate methods for sample size estimation. Only 63 (51.2%) reported exclusively appropriate methods for analysis.

These findings suggest that many investigators do not adequately attend to the methodological challenges inherent in group-randomized trials. These practices can lead to underpowered studies, to an inflated type 1 error rate, and to inferences that mislead readers. Investigators should work with biostatisticians or other methodologists familiar with these issues. Funders and editors should ensure careful methodological review of applications and manuscripts. Reviewers should ensure that studies are properly planned and analyzed. These steps are needed to improve the rigor and reproducibility of group-randomized trials.

The Office of Disease Prevention (ODP) at the National Institutes of Health (NIH) has taken several steps to address these issues. ODP offers an online course on the design and analysis of group-randomized trials. ODP is working to increase the number of methodologists who serve on grant review panels. ODP has developed standard language for the Application Guide and the Review Criteria to draw investigators' attention to these issues. Finally, ODP has created a new Research Methods Resources website to help investigators, reviewers, and NIH staff better understand these issues.

1. Introduction

Group-randomized trials, also called cluster-randomized trials, are comparative studies in which investigators randomize groups to study conditions, usually intervention and control, and observe members of those groups to assess the effects of the intervention (Campbell and Walters, 2014; Donner and Klar, 2000; Eldridge and Kerry, 2012; Hayes and Moulton, 2009; Murray, 1998). In this context, a group refers to any group that is not constituted at random, so that there is some connection among its members. For example, if worksites are

☆ Requests for reprints should be sent to David M. Murray, Ph.D., Associate Director for Prevention and Director of the Office of Disease Prevention, Division of Program Coordination Planning and Strategic Initiatives, Office of the Director, National Institutes of Health, 6100 Executive Boulevard, Suite 2B03, Bethesda, MD 20892.

* Corresponding author.
E-mail address: david.murray2@nih.gov (D.M. Murray).

[1] Note: the findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

https://doi.org/10.1016/j.ypmed.2018.03.010
Received 1 November 2017; Received in revised form 31 January 2018; Accepted 9 March 2018
Available online 16 March 2018
0091-7435/ Published by Elsevier Inc.

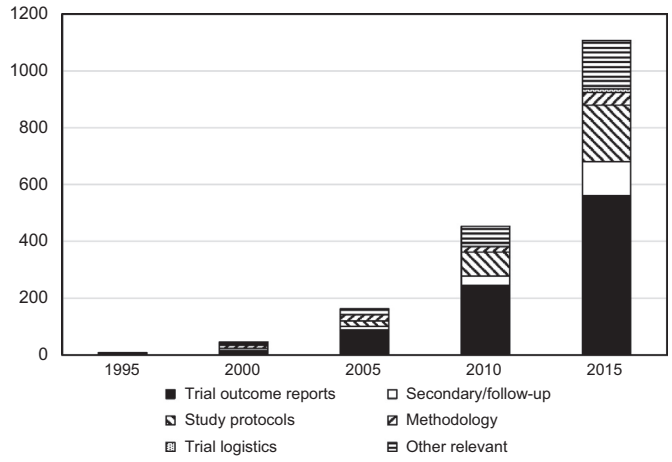

Fig. 1. PubMed titles or abstracts identifying group-randomized trials: 1995–2015[a].
[a]Based on a search of PubMed listings in the identified years for references to group- or cluster-randomized trials in the title or abstract. These figures are likely underestimates, as many papers that employ a group-randomized trial design do not provide such references in their title or abstract.
This work was performed in Bethesda Maryland during 2016–17.

randomized to study conditions and workers within those worksites are observed to assess the effects of an intervention, the worksites are the groups and the workers are the members.

Just as the randomized clinical trial is the gold standard in public health and medicine when allocation of individuals is possible, the group-randomized trial is the gold standard when allocation of groups is required (Murray, 1998). That will occur whenever investigators evaluate an intervention that operates at a group level, manipulates the social or physical environment, or cannot be delivered to individuals without substantial risk of contamination. These trials have become increasingly common over the last 20 years (Fig. 1); our search suggested a 280-fold increase in the number of group-randomized trials published in 2015 compared to 1995.

Turner et al. (2017a, 2017b) and Crespi (2016) recently reviewed the design and analytic challenges inherent in group-randomized trials. They note that the connections among group members create an expectation for positive intraclass correlation in observations taken on members of the same group (Kish, 1965); such correlation invalidates the independence assumption underlying the usual analytic methods and use of those methods will yield a Type I error rate that is inflated, often badly (Campbell and Walters, 2014; Cornfield, 1978; Donner and Klar, 2000; Eldridge and Kerry, 2012; Hayes and Moulton, 2009; Murray et al., 1998; Zucker, 1990). When only a few groups are randomized to each condition, the degrees of freedom (*df*) and power available for a valid test of the intervention effect will be limited. Finally, random assignment of only a few groups to each condition may jeopardize the internal validity of the trial by failing to distribute potential confounders evenly (Campbell and Walters, 2014; Donner and Klar, 2000; Eldridge and Kerry, 2012; Hayes and Moulton, 2009; Murray, 1998). Consideration must be given to these challenges as trials are planned and analyzed to support valid inference. Clear reporting is also important (Campbell et al., 2004).

Previous reviews have documented design and analytic problems in these trials (Brown et al., 2015; Crespi et al., 2011; Diaz-Ordaz et al., 2013; Diaz-Ordaz et al., 2014; Donner et al., 1990; Eldridge et al., 2008; Ivers et al., 2011; Murray et al., 2008; Rutterford et al., 2015; Simpson et al., 1995; Varnell et al., 2004). The most recent comprehensive review by Ivers et al. suggested that the methods had improved in trials published between 2000 and 2008; in particular, they reported that 61% and 70% of trials used appropriate methods for sample size and analysis, respectively (Ivers et al., 2011). In an earlier 2008 review focusing on cancer-related trials and covering much of the same time period, Murray et al. reported that only 24% and 45% of trials used appropriate methods for sample size and analysis, respectively (Murray et al., 2008). As a result, we have mixed evidence on whether the state of the practice with regard to the design and analysis of group-randomized trials has improved after 2000.

To the extent these problems continue, they contribute to the reproducibility challenges facing biomedical research (Collins and Tabak, 2014). To improve that situation, it is important to monitor the quality of the methods used and to encourage use of the best methods. This article assesses the state of the practice for group-randomized trials in studies published during 2011–2015 involving cancer-related risk factors and outcomes and offers recommendations for improvement.

2. Methods

The methods used for this review were based on those used in an earlier review by some of the same authors (Murray et al., 2008). We developed a list of groups used in these trials: (clinics, clusters, churches, colleges, communities, groups, hospitals, neighborhoods, physicians, practices, schools, units, wards, workplaces, worksites), hereafter represented as {groups}. We searched titles and abstracts in MEDLINE for human studies containing the following search term combinations: [cancer AND {groups}] AND [((community, cluster, group)(-,) (random*, rct)) OR ({groups}(were, were then, to be, are)(random*)) OR ((randomly assigned the {groups}) OR ({groups}(-based random*))]. We excluded articles based on the following key words in titles and abstracts: [(parallel)(-,)(group random*)] OR [(2-, 3-)(group random*)] OR [(two-, three-)(group random*)] OR [cluster random sampl*, rand* survey]. We also excluded articles based on key words in titles [protocol, review, metaanalysis, meta-analysis] and in publication types [review, meta-analysis]. The search identified 1451 candidate articles.

These articles were then manually inspected for the exclusion criteria and articles that met any of those criteria were excluded; some articles met more than one exclusion criteria. Articles reporting the results of studies in which groups were not randomly assigned to study conditions were excluded, as were studies that did not analyze observations taken on individual participants, and studies that lacked a clear statement that all groups were randomized to conditions. We excluded pilot studies because their goal is usually to evaluate intervention feasibility rather than efficacy. We excluded non-inferiority and equivalence trials because they are uncommon among group-randomized trials (Turner et al., 2017a) and cross-over and stepped-wedge designs because the impact of the intraclass correlation is reduced (Murray et al., 2010; Rhoda et al., 2011).

After these exclusions, we reviewed 123 primary articles (cf. Table A1) and 39 additional articles cited as background articles (cf. Table A2); these background articles were reviewed solely to inform the evaluation of methods for sample size estimation. Each article was reviewed independently by the first or second author and by two of the other six authors for design characteristics and methods used for sample size estimation and analysis of intervention effects.

For sample size estimation, we reviewed articles to determine whether authors reported evidence of taking group randomization into account a priori in establishing the size of the trial. Alternatives judged to be acceptable included reporting the expected intraclass correlation (Kish, 1987), coefficient of variation (Hayes and Moulton, 2009), or variance inflation factor (Donner et al., 1981), also known as the design effect (Kish, 1987).

For analysis of intervention effects, Table 1 (adapted from Murray et al. (2008)) presents the criteria used to judge whether methods were appropriate. Methods considered appropriate included mixed-model regression such as mixed model analysis of variance or covariance (ANOVA/ANCOVA) and linear and non-linear random coefficients models (Murray, 1998, 2001; Murray et al., 2004; Turner et al., 2017b); generalized estimating equations (GEE) (Liang and Zeger, 1986; Murray