Contemporary Clinical Trials 42 (2015) 41–50

Contents lists available at ScienceDirect

Contemporary Clinical Trials

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

Sample size calculations for the design of cluster randomized trials: A summary of methodology

Fei Gao [a,b,*], Arul Earnest [b,c], David B. Matchar [d,e], Michael J. Campbell [f], David Machin [f,g]

[a] National Heart Research Institute Singapore, National Heart Centre Singapore, 5 Hospital Drive, Singapore 169609
[b] Center for Quantitative Medicine, Duke-NUS Graduate Medical School, 8 College Road, Singapore 169857
[c] Department of Epidemiology and Preventive Medicine, Monash University, Level 6, The Alfred Centre, 99 Commercial Road, Melbourne, Victoria 3004, Australia
[d] Health Services & Systems Research, Duke-NUS Graduate Medical School, 8 College Road|, 169857, Singapore
[e] Department of Medicine (General Internal Medicine), Duke University Medical Center, 411 W. Chapel Hill Street, Suite 500, Durham, NC 27701, USA
[f] Medical Statistics Group, School of Health and Related Research, University of Sheffield, Regents Court, 30 Regent Street, Sheffield S1 4DA, UK
[g] Department of Cancer Studies and Molecular Medicine, Clinical Sciences Building, University of Leicester, Leicester Royal Infirmary, Leicester LE2 7LX, UK

ARTICLE INFO

Article history:
Received 9 December 2014
Received in revised form 26 February 2015
Accepted 28 February 2015
Available online 9 March 2015

Keywords:
Cluster randomized trial
Intra-class correlation
Sample size

ABSTRACT

Cluster randomized trial designs are growing in popularity in, for example, cardiovascular medicine research and other clinical areas and parallel statistical developments concerned with the design and analysis of these trials have been stimulated. Nevertheless, reviews suggest that design issues associated with cluster randomized trials are often poorly appreciated and there remain inadequacies in, for example, describing how the trial size is determined and the associated results are presented. In this paper, our aim is to provide pragmatic guidance for researchers on the methods of calculating sample sizes. We focus attention on designs with the primary purpose of comparing two interventions with respect to continuous, binary, ordered categorical, incidence rate and time-to-event outcome variables. Issues of aggregate and non-aggregate cluster trials, adjustment for variation in cluster size and the effect size are detailed. The problem of establishing the anticipated magnitude of between- and within-cluster variation to enable planning values of the intra-cluster correlation coefficient and the coefficient of variation are also described. Illustrative examples of calculations of trial sizes for each endpoint type are included.

1. Introduction

In contrast to clinical trials in which individual subjects are each randomized to receive one of the therapeutic options or interventions under test, the distinctive characteristic of a cluster trial is that specific groups or blocks of subjects (the clusters) are first identified and these units are assigned at random to the interventions. The term "cluster" in this context may be a household, school, clinic, care home or any other relevant grouping of individuals. When comparing the interventions in such *cluster randomized trials*, account must always be made of the particular cluster from which the data item is obtained.

A large and ever increasing number of cluster randomized trials have been conducted or are underway covering many aspects of cardiovascular related medicine. These include trials of cardiovascular guidelines [1], prescribing practice [2], community health awareness [3], breast feeding promotion on

* Correspondence to: F. Gao, National Heart Research Institute Singapore, National Heart Centre Singapore, 5 Hospital Drive, Singapore 169609. Tel.: +65 6704 2245; fax: +65 6844 9056.
E-mail addresses: gao.fei@nhcs.com.sg (F. Gao), arul_earnest@hotmail.com (A. Earnest), david.matchar@duke-nus.edu.sg (D.B. Matchar), m.j.campbell@sheffield.ac.uk (M.J. Campbell), dm113@leicester.ac.uk (D. Machin).

http://dx.doi.org/10.1016/j.cct.2015.02.011
1551-7144/

cardiometabolic risk factors in childhood [4], the effectiveness of a multifactorial intervention to improve both medication adherence and blood pressure control and to reduce cardiovascular events [5], and improving outcomes in patients with left ventricular systolic dysfunction [6]. In the Trial of Education And Compliance in Heart (TEACH) dysfunction trial [7], the clusters were the local pharmacists of patients with heart failure (HF) who had been hospitalised and then discharged into the community. The plan was that clusters were each randomized to one of the two interventions on a 1:1 basis: CONTROL or PHARM. Those pharmacists allocated PHARM would give their patients additional educational (motivational) support. Hence, all the patients within a particular cluster received the same intervention. A patient experiencing any one of a readmission, emergency room visit or mortality due to HF was regarded as a failure.

There are numerous publications describing design, analysis and reporting issues concerned with cluster randomized trials, including text books [8–11], and the associated challenges [12]. However much of the literature is fragmented and some quite old (though still relevant). Further some of the articles are quite technical in nature so investigators may find it difficult to determine best practice. A review of cluster trials [13], published subsequent to the 2004 extension of the CONSORT guidelines [14,15], concluded that the methodological quality of cluster trials often remains suboptimal.

To facilitate and improve this situation, we focus on methods of determining the number of subjects (and clusters) required with the aim to provide a compact but comprehensive reference for those designing cluster trials.

2. General design considerations

2.1. Individually randomized trials — continuous outcome measure

At the close of a clinical trial, and once all the data collection is complete, a comparison will be made between the interventions with respect to the primary endpoint. For the case of two interventions, *Standard* (S) and *Test* (T), with n_S and n_T patients respectively randomized *individually* to each, the statistical process for a continuous outcome measure, y, is made by comparing the corresponding means $\bar{y}_S$ and $\bar{y}_T$ by use of Student's t-test. This tests the null hypothesis that the difference $\delta = \mu_T - \mu_S = 0$, where μ_S and μ_T are the true or population means of interventions S and T. If the null hypothesis is rejected then we conclude that μ_S and μ_T differ.

However, prior to this analysis, the trial must first be designed and conducted. In general, critical decisions to be made by the design team are the choice (and number) of interventions to compare and the endpoint measure which will be used for the evaluation. A vital detail is the difference in the outcome (the *effect size* or δ_{Plan}) between the randomized interventions which might be anticipated. Such a difference should be one (if established) of sufficient clinical importance to justify the *expense* of conducting the planned trial and likely to lead to changes in clinical practice. Also required is the standard deviation (SD), σ_{Plan}, of the endpoint variable of concern. A further design option is the choice of the ratio of subjects $1{:}\varphi$ allocated to S and T respectively (see below).

Once these aspects are provided, the numbers of subjects to be randomized to each intervention for a continuous endpoint is [16]:

$$n_S = \left(\frac{1+\varphi}{\varphi}\right)\frac{\left(z_{1-\alpha/2}+z_{1-\beta}\right)^2}{\left(\delta_{Plan}/\sigma_{Plan}\right)^2} + \left[\frac{z^2_{1-\alpha/2}}{2(1+\varphi)}\right], n_T = \varphi n_S \qquad (1)$$

giving a total $N = n_S + n_T$.

Here, α and β are the Type I and Type II errors and $1-\beta$ is the *power*.

Further, $z_{1-\alpha/2}$ and $z_{1-\beta}$ are values with probabilities of $\alpha/2$ and β respectively in the upper tail of the standard Normal distribution. Typically $\alpha = 0.05$ leading to $z_{1-0.05/2} = z_{0.975} = 1.9600$ while $\beta = 0.2$ or 0.1 leading to $z_{0.8} = 0.8416$ and $z_{0.9} = 1.2816$ respectively.

The final term $\left[\frac{z^2_{1-\alpha/2}}{2(1+\varphi)}\right]$ in Eq. (1) applies only when the sample size is small. However, when $\alpha = 0.05$ and $\varphi = 1$, this implies adding $\left[\frac{1.96^2}{2\times(1+1)}\right] = \frac{3.8416}{4} \approx 1$ unit extra to each intervention group.

An alternative is first to assume $\varphi = 1$ in Eq. (1), to obtain n subjects for each intervention and then calculate the final numbers per intervention using

$$n_S = \frac{n(1+\varphi)}{2\varphi}, n_T = \frac{n(1+\varphi)}{2}. \qquad (2)$$

This increases the initial total number of subjects N from $2n$ to $\frac{n(1+\varphi)^2}{2\varphi}$ which, if $\varphi = 0.5$, implies that $N = 2.25n$. If, as we will be concerned with later, it is the number of clusters that is being calculated then k, k_S, k_T and K replace the corresponding n's.

Community based exercise programme [17]

In this trial, 8 clusters for *Control* and 4 for *Test* were used for evaluating the programme as budgetary constraints limited the number of *Test* facilities (clusters) available whereas: "… the relative costs of including controls were very small …". Thus instead of using a 1:1 design, with 4 clusters, per intervention, a 2:1 allocation using 12 clusters enabled a larger trial with greater power to be conducted without increasing the number of T clusters, k_T.

2.2. Cluster randomized trials

When the randomized allocation applies to the clusters, the basic principles for sample size calculation still apply although modifications are required. To illustrate these we first describe