

Randomizing patients by family practice: sample size estimation, intracluster correlation and data analysis

Roxanne H Cosby^a, Michelle Howard^a, Janusz Kaczorowski^{a,b,c}, Andrew R Willan^{b,c} and John W Sellors^{a,b}

Cosby RH, Howard M, Kaczorowski J, Willan AR and Sellors JW. Randomizing patients by family practice: sample size estimation, intracluster correlation and data analysis. *Family Practice* 2003; **20**: 77–82.

Background. Cluster randomized controlled trials increasingly are used to evaluate health interventions where patients are nested within larger clusters such as practices, hospitals or communities. Patients within a cluster may be similar to each other relative to patients in other clusters on key variables; therefore, sample size calculations and analyses of results require special statistical methods.

Objective. The purpose of this study was to illustrate the calculations used for sample size estimation and data analysis and to provide estimates of the intraclass correlation coefficients (ICCs) for several variables using data from the Seniors Medication Assessment Research Trial (SMART), a community-based trial of pharmacists consulting to family physicians to optimize the drug therapy of older patients.

Methods. The study was a paired cluster randomized trial, where the family physician's practice was the cluster. The sample size calculation was based on a hypothesized reduction of 15% in mean daily units of medication in the intervention group compared with the control group, using an alpha of 0.05 (one-tailed) with 80% power, and an ICC from pilot data of 0.08. ICCs were estimated from the data for several variables. The analyses comparing the two groups used a random effects model for a meta-analysis over pairs.

Results. The design effect due to clustering was 2.12, resulting in an inflation in sample size from 340 patients required using individual randomization, to 720 patients using randomization of practices, with 15 patients from each of 48 practices. ICCs for medication use, health care utilization and general health were <0.1; however, the ICC for mean systolic blood pressure over the trial period was 0.199.

Conclusions. Compared with individual randomization, cluster randomization may substantially increase the sample size required to maintain adequate statistical power. The differences in ICCs among potential outcome variables reinforce the need for valid estimates to ensure proper study design.

Keywords. Cluster randomization, intracluster correlation, primary care, RCT.

Introduction

The randomized controlled trial (RCT) is seen as the gold standard of evidence in evaluating health service

interventions.¹ Many interventions are directed at providers, with the intent of changing patient outcomes. This is one reason why cluster randomized trials are used increasingly in primary care health services research.² The main feature of such interventions is that patients are nested within larger clusters such as physician practices, hospitals or communities, and that the intervention is applied at that level, while the outcomes are measured at the patient level. This study design necessitates special statistical considerations for sample size estimation and data analysis.³

Often, patients within the same practices or clusters are more similar to each other, with respect to key

Received 4 February 2002; Revised 16 July 2002; Accepted 9 September 2002.

Departments of ^aFamily Medicine and ^bClinical Epidemiology and Biostatistics, McMaster University and ^cCentre for the Evaluation of Medicines, St. Joseph's Healthcare, Hamilton, Ontario, Canada. Correspondence to Roxanne Cosby, Hamilton Health Sciences Corporation, Henderson Site, 699 Concession St. Level 3-62, Hamilton, ON L8V 5C2, Canada; E-mail: cosbyr@mcmaster.ca

confounders or the outcomes of interest, than to patients in other practices. For example, the management of patients with a given condition by the same physician is more likely to be similar than management of patients with the same condition by different physicians. The end result of using a cluster RCT is that the design is not as statistically efficient as a patient RCT. This is because greater homogeneity of members in the clusters increases the standard error of the estimate of the treatment effect,⁴ resulting in a loss of power to detect a difference between the intervention and control groups. Therefore, a compensatory inflation of sample size is required to maintain power in a cluster randomized trial. The intracluster correlation coefficient (ICC) is used in the calculation of the inflation factor, or design effect. The sample size is calculated assuming randomization of individuals, and then increased by the amount indicated by the inflation factor.³ The ICC is defined as the ratio of between-cluster variance to total variance. As the magnitude of the ICC increases, the more individuals within clusters resemble one another and the less they resemble those from other clusters with respect to the variable of interest. An ICC of zero indicates that people within and among the clusters are completely independent with respect to that variable. In contrast, an ICC of 1.0 indicates that people within a cluster are identical, but each cluster differs from the others. As an ICC increases, the sample size required to detect a significant difference for that variable increases.

To address this issue in cluster randomized trials, investigators need valid estimates of the ICC for their outcomes of interest which, together with the number of observations per cluster or practice, determine the magnitude of the necessary inflation factor for the sample size in the cluster randomized design.⁵

Cluster randomized trials also require special considerations for data analysis. Researchers often fail to take into account the clustered nature of their data, and mistakenly perform their statistical analyses at the patient level, when the physician or practice was the unit of randomization.⁶ This ‘unit of analysis’ error results in an increased probability of rejecting the null hypothesis.⁷

The objectives of the Seniors Medication Assessment Research Trial (SMART) were to compare over 5 months the medication outcomes, management of hypertension, hyperlipidaemia and diabetes, health care utilization and costs, and health-related quality of life among community-dwelling seniors (aged 65 years and older) whose physicians received the pharmacist consultation intervention, compared with those whose physicians did not receive the intervention.

This paper illustrates the calculations used to estimate the sample size and the statistical approach used to analyse the data in our paired cluster randomized trial. Estimates of ICCs from the trial results for some common outcomes, which may be useful to estimate sample sizes in future trials in the primary care setting, are also presented.

Methods

The design of our trial entailed an intervention directed at physicians, with outcomes measured primarily at the patient level. The recruitment methods have been described previously.⁸ Briefly, the study involved 48 randomly selected family physicians (including GPs) in Southern Ontario (69.6% participation rate) who were randomized to either the control or the intervention group to test the effectiveness of community pharmacists consulting to family physicians. Two physicians (one pair) were recruited within each of 24 geographic areas, defined as the first three digits of the postal code of a corresponding pharmacist’s practice address, and then, within each pair, randomly allocated to the pharmacist intervention or control group. Paired randomization was used in order to increase the comparability between study seniors on potential covariates, such as socio-demographic factors and access to health care resources. This type of randomization is only applicable to paired cluster designs. Otherwise, randomization approaches used for non-clustered designs such as stratification, block randomization or minimization can be used. In each practice, 15–20 seniors taking five or more medications daily according to their chart were selected randomly (69.5% participation rate) to participate in the trial. The primary outcomes were assessed on each patient, but the intervention itself was aimed at the physician.

The seniors were asked to bring all of their regular medications to a baseline interview with a study nurse, prior to randomization of the family practices. A regular medication was defined as one that was taken in the last 2 days. Drug name, strength, units per dose and times taken per day were recorded for each medication. One unit of medication was defined as one tablet, one teaspoon of liquid, one drop (for eye drops) or one application of cream/ointment. The health conditions of the seniors were captured by medical chart audits prior to randomization of the practices, and were verified by the family physician and coded systematically by a physician investigator (JS) using ICD-9 coding. For those with hypertension, hyperlipidaemia or diabetes, information was collected from the charts on blood pressure, cholesterol and blood sugar readings, and the occurrence of complications such as renal failure, over 5 months after randomization.

Details of the intervention have been described elsewhere.⁹ The pharmacist assigned to each intervention practice met with each participating senior in the practice to review their medications, and then formulated written recommendations for the physician and discussed them in a face-to-face meeting. The recommendations involved the identification of drug-related problems and suggestions for optimizing each patient’s drug regimen.

We hypothesized that the pharmacist intervention should have been capable of decreasing the average number of units of medication taken daily by 15%. We

felt that this was a clinically meaningful difference, and this was borne out by our pilot study.¹⁰ It represents a reduction of ~2 daily units of medication, which should translate into a reduction of the complexity of the medication regimen, and a reduction in drug costs. The mean number of units of daily medications (14.7) and standard deviation (8.1) from the pilot study were used in the sample size calculation. A one-tailed test was used because our hypothesis was unidirectional. The sample size calculation used was described by Donner *et al.*³ The ICC for the number of daily units of medication taken was estimated to be 0.08, based on the pilot study. This ICC was used to calculate the inflation factor, or design effect,⁵ using the following formula:

$$\text{Inflation factor} = 1 + (\text{number of patients per practice} - 1) \times \text{ICC}$$

To account for the paired cluster design of the trial, the analyses that compared the intervention and control arm or groups were performed using the methods proposed by Thompson *et al.*,¹¹ which uses a random effects model for a meta-analysis over pairs. The mean difference on the outcome variable between groups was a weighted average of the mean differences across clusters (d_j). The formula for the weighted mean difference given by Thompson is:

$$(\text{mean } d_w) = \sum_j w_j d_j / \sum_j w_j$$

where $w_j = 1/\text{var}(d_j)$. Variance for the weighted mean is expressed as:

$$\text{var}(\text{mean } d_w) = 1 / \sum_j w_j$$

Data from the SMART results were used to calculate ICCs for several outcome variables. ICCs were calculated from the *F* statistics of one-way ANOVA analyses using the formula:

$$s_{\text{between}}^2 / (s_{\text{between}}^2 + s_{\text{within}}^2)$$

where s^2 = variance.

Results

Sample size estimation

For the sample size estimation of SMART, the design effect was calculated to be 2.12 based on 15 seniors from each of 48 family practices. The number of patients recruited per practice was increased to 20, where possible, to account for losses to follow-up. If the number of physicians or the number of patients per practice had been altered, this would have resulted in higher or lower inflation factors as shown in Table 1. For example, for any given value of ICC, increasing the cluster size results in the need for fewer physician practices in the trial. When deciding on sample size, pragmatism often becomes the overriding issue. Increasing the cluster size may make it difficult to find enough patients in any given

physician practice that meet the inclusion criteria. Conversely, decreasing the cluster size and increasing the number of clusters needed also has ramifications in the form of making the trial logistically more difficult and increasing costs (we would have required more expanded role pharmacists). Thus, we strove to strike a balance between all of these relevant considerations. Table 1 also shows the effects on sample size of varying the ICC for the main outcome measure, number of daily units of medication. The sample size required using the assumption of individual randomization was calculated to be 340 using a standard formula.¹²

Sample description

The estimated median practice size of the 48 physicians was 2142 patients (interquartile range = 1300 patients). Among the 48 practices, 10 663 patients were aged 65 years and older. After reviewing the charts to determine initial eligibility, 2078 (19.5%) seniors were found to be taking ≥ 5 daily medications.

The average age of the 889 seniors participating in the study was 74.0 years (SD = 6.1 years), and 62.8% (558/889) were female. The average number of medications taken daily including prescription and over-the-counter was 8.1 (SD = 3.4). The five most common health conditions among the seniors were hypertension (55.0%; 489/889), osteoarthritis (47.4%; 421/889) ischaemic heart disease (37.0%, 329/889), hyperlipidaemia (32.5%; 289/889) and angina (23.7%; 211/889).

Paired cluster analysis

Figure 1 depicts a graphical example of the data analysis methods using the main outcome, mean daily units of

TABLE 1 Total number of seniors (and physicians) required to detect a 15% decrease in daily units of medication (from 14.7 to 12.5, SD = 8.1 based on pilot study), given ICC = 0.08, 0.10 or 0.15, alpha = 0.05 (one-tailed) and power = 0.80

Cluster size (no. of seniors)	Intracluster correlation coefficient		
	0.08	0.10	0.15
10	580 (58 practices) (IF = 1.72)	640 (64 practices) (IF = 1.90)	800 (80 practices) (IF = 2.35)
15	720 (48 practices) (IF = 2.12)	810 (54 practices) (IF = 2.40)	1080 (72 practices) (IF = 3.10)
20	840 (42 practices) (IF = 2.52)	980 (49 practices) (IF = 2.90)	1300 (65 practices) (IF = 3.85)
Using individual randomization	340		

The number of seniors was increased to the nearest multiple of 10 to be divisible by 24 physicians.
IF, Inflation factor.

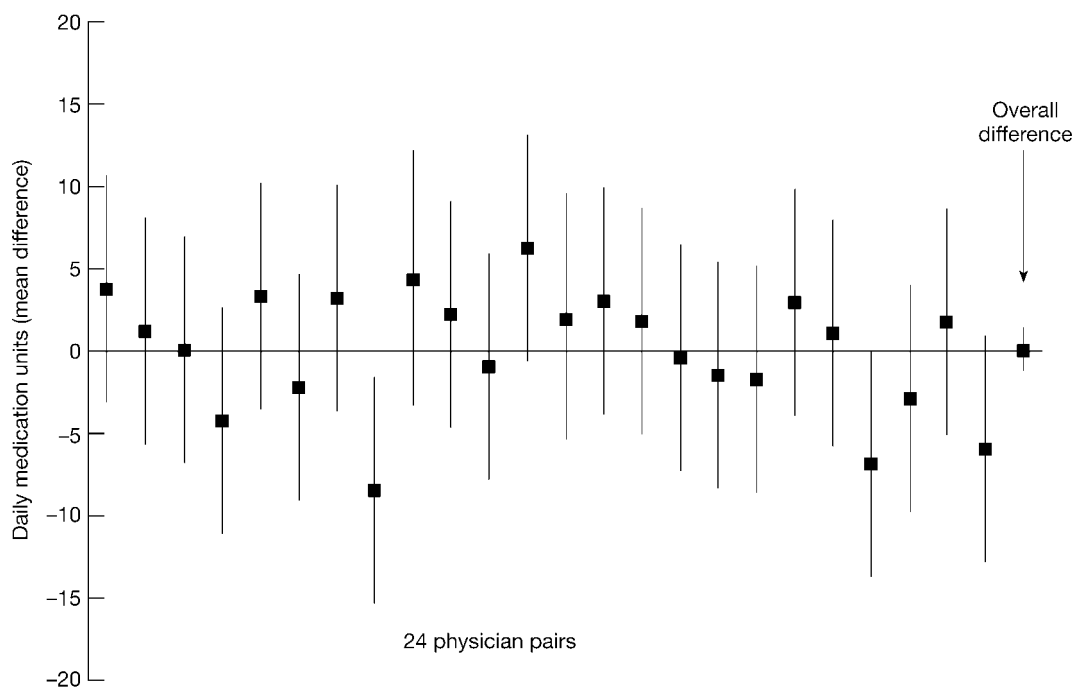


FIGURE 1 The mean differences and 95% CIs for the daily units of medication in the control and intervention paired practices at the end of the SMART (difference = intervention practice – control practice)

medication. The mean difference between groups and the 95% confidence intervals (CIs) are plotted for the 24 physicians pairs as well as the weighted overall mean difference with the 95% CI. There are several methodological alternatives that can be used to analyse clustered data, and they include the unweighted paired *t*-test in which a summary statistic is calculated for each cluster^{2,13} and multi-level models.¹⁴ One can also use non-parametric tests such as Wilcoxon's signed rank test.¹⁵

Estimates of ICCs

Table 2 provides estimates of ICCs for outcomes that could be relevant to primary care studies in the elderly. The ICC calculated from the baseline data of SMART for daily medication units was actually 0.06, thus we slightly overcompensated for the effects of clustering based on this main outcome variable. The highest ICC (0.199) was found for mean systolic blood pressure during the trial. The lowest ICC (0.0000035) was found for the occurrence of a renal or cardiovascular outcome or death among seniors with hypertension, hyperlipidaemia or diabetes mellitus during the 5 months after randomization, and the 95% CI included 0.

Discussion

The recruitment of 20 seniors from each of 48 physicians in the SMART study was feasible and ensured adequate power for the analyses. In designing a cluster randomized trial, there is a trade-off between the number of

observations recruited per cluster, and the number of clusters to be used. As the natural cluster size becomes larger, the ICC generally declines.¹⁶ However, it often is desirable to increase the number of clusters rather than the number of observations within each cluster, since recruiting more observations from already large clusters will yield minimal increases in statistical power.¹⁷ The inflation factor (and hence the total sample size) in the SMART study could have been reduced by increasing the number of physicians participating in the trial, but this would have been very difficult to do.

Our analysis methods, which took into account the paired clustered nature of the data, ensured that variance was not underestimated, by taking into account both between-cluster and between-individual variances. Although there was considerable variability in the mean differences of daily medication units between the 24 pairs of control and intervention group practices, the mean overall difference was very small and not statistically significant 5 months after randomization.

The results of this study are based on people aged 65 years and older taking multiple medications and may not be generalizable to all seniors. Our measures of ICC for the outcomes in Table 2 probably reflect this select population, which was chosen based on the complexity of seniors' medication regimens.

Compared with other ICCs, the ICC for systolic blood pressure was high (0.199), at least twice the magnitude of any other ICC. This suggests that seniors within practices were much more similar to each other than to seniors in other practices. This may in part reflect the hypertension

TABLE 2 The ICCs for outcome measures using the baseline data from the cluster randomized trial of community pharmacists consulting to family physicians (n = 889 unless otherwise specified)

Variable	Mean, SD	ICC (95% CI)
Total daily number of drugs	8.1, 3.4	0.0906 (0.08812–0.09308)
Total daily units ^a of drugs	12.4, 9.1	0.0636 (0.06180–0.06540)
Total daily dosage of drugs	10.8, 5.9	0.0496 (0.04817–0.05104)
Total daily drug cost (\$CDN)	5.07, 6.09	0.0144 (0.01395–0.01486)
Hospitalization over 5 months	9.2 (82/889)	0.0019 (0.0018–0.0020)
No. of physician visits over 5 months	4.8, 3.7	0.0697 (0.0677–0.0717)
Emergency department visit over 5 months	12.9 (115/889)	0.0206 (0.0199–0.0212)
Mean systolic blood pressure (n = 448)	140.1, 16.1	0.1990 (0.1942–0.2038)
Mean total cholesterol (n = 289)	5.2, 0.9	0.0480 (0.0465–0.0495)
Mean fasting plasma glucose (n = 165)	8.2, 2.9	0.0207 (0.0199–0.0215)
% with any cardio-renal outcome (n = 567)	5.9 (30/607)	0.0000035 (–0.00005–0.00006)
SF-36 ²¹ Question 1 ^b	3.4, 1.1	0.0648 (0.0630–0.0666)
SF-36 Physical Component Scale	40.8, 11.0	0.0207 (0.0201–0.0213)
SF-36 Mental Component Scale	66.8, 10.0	0.0245 (0.0238–0.0252)
% taking a known inappropriate drug for the elderly ²²	22.6 (201/889)	0.0296 (0.0287–0.0305)
% taking a known inappropriate combination of medications in the elderly ²³	11.6 (103/889)	0.0804 (0.0782–0.0826)

^a Defined as one tablet, one teaspoon of liquid, one drop (for eye drops) or one application of cream/ointment.

^b Health rating (5-point scale, 1 = excellent, 5 = very poor).

management style of physicians, with respect to measurement or pharmacotherapy. In a study of preventive health practices in Canada, Baskerville *et al.*¹⁸ found an ICC of 0.18 for evidence-based hypertension treatment among adult patients of group practices. These researchers also found a high ICC for other behaviours that reflect the management style of physicians, including ordering a chest X-ray for smokers (ICC = 0.66) and smoking cessation counselling (ICC = 0.23).¹⁸ In the UK, ICCs for the proportion of patients with controlled hypertension in 18 general practices ranged from 0.05 to 0.06 using different hypertension management guidelines to define control.¹⁹ In contrast, in a community intervention study where clusters were cities, the ICC for systolic blood pressure was only 0.01.²⁰ Health-related outcomes that are managed by physicians would be expected to have much more homogeneity within physician practices compared with within communities.

Similarly to other studies, we found that ICCs for variables less related to physician management, such as number of visits to the emergency room and self-reported health, tended to be lower, whereas ICCs for process outcomes, or physician management, are often higher than those for patient outcomes.⁴ The management strategies of a particular condition within a physician's practice would be expected to be more homogeneous than the patients' outcomes or responses with respect to

that management. Campbell *et al.* calculated ICCs for several cluster randomized trials in primary care settings.⁴ ICCs for process outcomes in physician practices such as number and appropriateness of referrals ranged from 0.01 to 0.24, while outcome measures among patients for the SF-36 ranged from 0.007 to 0.01. They also found that the ICCs for process outcomes in the secondary care setting of hospitals were higher than those in primary care settings. Physicians' practices within a particular institution are likely to be more homogeneous than in primary care settings, where physicians practice more independently.

The use of appropriate study designs and analytic techniques will improve further the quality of evidence obtained from research in primary care. Some authors have published estimates of ICCs on the Internet,⁴ and we urge other researchers to disseminate their results.

The description of this study of seniors taking multiple medications, and the publication of other ICCs from a wide range of content areas and settings, hopefully will enable primary care researchers to surmount these methodological issues.

References

- Sibbald B, Roland M. Understanding controlled trials: why are randomised controlled trials important? *Br Med J* 1998; **316**: 201.

- ² Campbell MK, Mollison J, Steen N, Grimshaw JM, Eccles M. Analysis of cluster randomized trials in primary care: a practical approach. *Fam Pract* 2000; **17**: 192–196.
- ³ Donner A, Birkett N, Buck C. Randomization by cluster, sample size requirements and analysis. *Am J Epidemiol* 1981; **114**: 906–915.
- ⁴ Campbell M, Grimshaw J, Steen N. Sample size calculations for cluster randomised trials. *J Health Serv Res Policy* 2000; **15**: 12–16.
- ⁵ Kerry SM, Bland JM. The intracluster correlation in cluster randomization. *Br Med J* 1998; **316**: 1455.
- ⁶ 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.
- ⁷ Kerry SM, Bland JM. Trials which randomize practices 1: how should they be analysed? *Fam Pract* 1998; **15**: 80–83.
- ⁸ Sellors J, Cosby R, Trim K *et al*. Recruiting family physicians and patients for a clinical trial: lessons learned. *Fam Pract* 2002; **19**: 99–104.
- ⁹ Sellors J, Sellors C, Woodward C *et al*. Expanded role pharmacists consulting in family physicians' offices—a highly acceptable program model. *Can Pharm J* 2001; **134**: 27–31.
- ¹⁰ Sellors C, Dalby DM, Howard M, Kaczorowski J, Sellors J. A pharmacist consultation service in community-based family practices: a randomized controlled trial in seniors. *J Pharm Technol* 2002; **17**: 264–269.
- ¹¹ Thompson SG, Pyke SDM, Hardy RJ. The design and analysis of paired cluster randomized trials: an application of meta-analysis techniques. *Stat Med* 1997; **16**: 2063–2079.
- ¹² Rosner B. Hypothesis testing: two-sample inference. In Rosner B. *Fundamentals of Biostatistics*, 4th edn. Belmont (CA): Wadsworth Publishing Company, 1995: 251–298.
- ¹³ Wears RL. Advanced statistics: statistical methods for analyzing cluster and cluster-randomized data. *Acad Emerg Med* 2002; **9**: 330.
- ¹⁴ Rice N, Leyland A. Multilevel models: applications to health data. *J Health Serv Res Policy* 1996; **1**: 154.
- ¹⁵ Lehman EL. *Non-parametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day, 1975.
- ¹⁶ Donner A. An empirical study of cluster randomization. *Int J Epidemiol* 1982; **11**: 283–286.
- ¹⁷ Diwan VK, Eriksson B, Sterky G, Tomson G. Randomization by group in studying the effect of drug information in primary care. *Int J Epidemiol* 1992; **21**: 124–130.
- ¹⁸ Baskerville NB, Hogg W, Lemelin J. The effect of cluster randomization on sample size in prevention research. *J Fam Pract* 2001; **50**: 241–246.
- ¹⁹ Fahey TP, Peters TJ. What constitutes controlled hypertension? Patient based comparison of hypertension guidelines. *Br Med J* 1996; **13**: 93–96.
- ²⁰ Hannan PJ, Murray DM, Jacobs DR Jr, McGovern PG. Parameters to aid in the design and analysis of community trials: intraclass correlations from the Minnesota Heart Health Program. *Epidemiology* 1994; **5**: 88–95.
- ²¹ McHorney C, Ware JJ, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993; **31**: 247–263.
- ²² Beers MH. Explicit criteria for determining potentially inappropriate medication use by the elderly. *Arch Intern Med* 1997; **157**: 1531–1536.
- ²³ Tamblyn RM, McLeod PJ, Abrahamowicz M *et al*. Questionable prescribing for elderly patients in Quebec. *Can Med Assoc J* 1994; **150**: 1801–1809.