



Published in final edited form as:

*Contemp Clin Trials*. 2009 March ; 30(2): 129–132. doi:10.1016/j.cct.2008.10.002.

## An Application of a Modified Constrained Randomization Process to a Practice-Based Cluster Randomized Trial to Improve Colorectal Cancer Screening

Paul J Nietert, PhD<sup>1</sup>, Ruth G Jenkins, PhD<sup>2</sup>, Lynne S Nemeth, PhD, RN<sup>3</sup>, and Steven M Ornstein, MD<sup>2</sup>

<sup>1</sup> Department of Biostatistics, Bioinformatics, and Epidemiology, University of South Carolina, Charleston, SC

<sup>2</sup> Department of Family Medicine, University of South Carolina, Charleston, SC

<sup>3</sup> College of Nursing Medical, University of South Carolina, Charleston, SC

### Abstract

**Background**—When designing cluster randomized trials, it is important for researchers to be familiar with strategies to achieve valid study designs given limited resources. Constrained randomization is a technique to help ensure balance on pre-specified baseline covariates.

**Methods**—The goal was to develop a randomization scheme that balanced 16 intervention and 16 control practices with respect to 7 factors that may influence improvement in study outcomes during a 4-year cluster randomized trial to improve colorectal cancer screening within a primary care practice-based research network. We used a novel approach that included simulating 30,000 randomization schemes, removing duplicates, identifying which schemes were sufficiently balanced, and randomly selecting one scheme for use in the trial. For a given factor, balance was considered achieved when the frequency of each factor's sub-classifications differed by no more than 1 between intervention and control groups. The population being studied includes approximately 32 primary care practices located in 19 states within the U.S. that care for approximately 56,000 patients at least 50 years old.

**Results**—Of 29,782 unique simulated randomization schemes, 116 were determined to be balanced according to pre-specified criteria for all 7 baseline covariates. The final randomization scheme was randomly selected from these 116 acceptable schemes.

**Conclusions**—Using this technique, we were successfully able to find a randomization scheme that allocated 32 primary care practices into intervention and control groups in a way that preserved balance across 7 baseline covariates. This process may be a useful tool for ensuring covariate balance within moderately large cluster randomized trials.

---

Please address all correspondence to: Paul J Nietert, PhD, Associate Professor, Department of Biostatistics, Bioinformatics, and Epidemiology, Medical University of South Carolina, 135 Cannon Street, Suite 303, PO Box 250835, Charleston, SC 29425, E-mail: E-mail: nieterpj@musc.edu, Phone: 843-876-1204, Fax: 843-876-1126.

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

## Keywords

Randomization techniques; cluster randomized trials; covariate balance; study design; practice based research networks; colorectal cancer screening

---

## Introduction

“Cluster” (or “group”) randomized trials (CRTs), are trials that use a cluster as the unit of randomization and cluster members as the units of analysis [1]. In contrast to traditional randomized trials in which patients are randomized, CRTs are often used in circumstances when the intervention is administered at the level of the cluster.

The use of CRTs has been steadily increasing since 1980 [2]. Within primary care research, these study designs might be becoming more popular for several reasons [3], including a greater national emphasis on improving quality of care (often requiring systems-level interventions) [4], the development and growth of practice based research networks (which often provide excellent infrastructure for CRTs) [5], and the continued adoption of electronic health records (which greatly facilitate data collection) [6]. As CRTs become more common, it is important that researchers become more familiar with the ways in which they differ from traditional randomized clinical trials.

Because CRTs often include a small number of groups relative to the number of subjects, the chance of treatment groups being imbalanced on important relevant covariates may be unacceptably high [7–9]. Should the clusters be randomized in an imbalanced fashion, it may be difficult to determine whether the estimated treatment effect is influenced by baseline imbalance. While it may be possible to control for such imbalance in the analysis phase, results may still be deemed “suspect” [10].

Several mechanisms have already been proposed for balancing baseline covariates across treatment groups. The processes, collectively referred to as “constrained randomization”, have been discussed in the statistical literature since the 1970’s [11,12]. These methods include stratifying clusters prior to randomization [13,14] or pairwise matching of clusters [15,16]. Such techniques are ideal when there are sufficient numbers of subjects for each stratum. However, having large numbers of strata in studies with small numbers of randomization units can be impractical because of sparseness of subjects within each stratum [12]. Additionally, imbalance may be minimized by selecting an appropriately balanced randomization scheme from all possible allocations of clusters to treatments [17,18]. However, as the number of clusters to be randomized increases, the number of possible randomization schemes grows extremely fast, and enumerating all possible randomization schemes may be impractical due to computational limitations. In this paper we propose an alternative method to achieve balance across several ( $> 2$ ) covariates in a CRT.

The research context is a 4-year CRT, called Colorectal Cancer (CRC) Screening in Primary Care Practice (C-TRIP). The population being studied includes 32 primary care practices located in 19 states within the U.S. that care for approximately 56,000 patients at least 50 years old. They share a common electronic medical record system and participate in a practice based research network called Practice Partner Research Network (PPRNet).

The study protocol required 16 practices to be randomized to receive a specific practice-focused intervention based upon the PPRNet quality improvement model [19,20] for improving CRC screening and 16 practices to be randomized to a usual care control group. The multi-faceted intervention incorporates quarterly provider feedback reports, semi-annual site visits by

research investigators to help identify strategies for improvement, and annual network meetings. The primary outcome is the proportion of active patients 50 year of age or older up-to-date with any form of CRC screening.

Study investigators desired for the intervention and control groups to be balanced on 7 factors of interest that might influence the degree of improvement in CRC screening. Factors included baseline CRC screening performance (3 tertiles), presence/absence of state mandated insurance coverage for CRC screening, past network experience with quality improvement projects (yes/no), practice specialty (internal medicine or family medicine), number of healthcare providers (3 groupings: 1 or 2, 3 or 4, 5 or more), geographic region (East, South, Midwest, West), and whether or not the practice is a residency training program.

## Methods

Because there were over 600 million ways to randomize these 32 practices into 2 groups, enumerating all possible randomization schemes and assessing their level of covariate balance was computationally impractical. We modified this approach by significantly reducing the number of randomization schemes being assessed. This process included simulating randomization schemes (i.e. randomly assigning the 32 practices into 1 of the 2 treatment groups, 16 practices per group), removing duplicate schemes, identifying which ones were sufficiently marginally balanced on the pre-specified covariates of interest (defined below), and randomly selecting one scheme from all possible balanced schemes, with each having an equal probability of being selected.

Based upon preliminary timings of our algorithm, we determined that it would be possible to generate and assess approximately 30,000 randomization schemes within a reasonable amount of time (24-hours). The computer was then programmed to conduct 30,000 simulations, to find 1 or more sufficiently balanced randomization schemes. Were a different study to have fewer pre-specified covariates, more potential randomization schemes could be generated with a 24-hour window, since balance assessment would be quicker.

To create a randomization scheme, we allocated 32 practices into 2 treatment groups by assigning a random number to each practice using a statistical software program (i.e. SAS v9.1.3, Cary NC), sorting the practices by the random number, selecting the first 16 practices to be in the intervention group, and selecting the last 16 practices to be in the intervention group, in effect randomly assigning each practice to intervention or control. This process was repeated to generate 30,000 randomization schemes. Quicker algorithms may exist, but this method is generalizable to other software packages.

To assess balance for a simulated scheme, we generated 2-way cross-tabulations of treatment group by each factor of interest and saved the frequencies in a separate dataset, 1 observation per scheme. For a given factor, balance was considered to have been achieved when the frequency count of practices within each factor's sub-classifications differed by no more than 1 between intervention and control groups. For example, since there were 8 practices in states with mandated insurance coverage for CRC screening, 4 were forced to be in the intervention group and 4 had to be in the control group; a 5:3 split was considered imbalanced. A series of "If-Then" statements were used to determine whether marginal balance was achieved for all of the 7 factors. Once the final randomization scheme was selected, we also compared intervention and treatment groups on other baseline variables using non-parametric Wilcoxon rank sum tests.

## Results

Of the 30,000 generated randomization schemes, 228 were duplicates, assigning the same practices to the treatment and control groups. Of 29,782 uniquely simulated randomization schemes, 116 (0.39%) were balanced on all baseline factors of interest. The final randomization scheme was randomly selected from these 116 acceptable schemes. By design, none of the 7 relevant baseline factors differed significantly between intervention and control groups (see Table 1). Additionally, intervention group practices did not differ significantly from control group practices with respect to number of eligible patients ( $p=0.99$ ), baseline percent of eligible patients up-to-date with CRC screening ( $p=0.87$ ), average age among eligible patients ( $p=0.84$ ), and percent of eligible patients who are male ( $p=0.84$ ).

## Discussion

There is a growing need to optimize designs for CRTs. We have described a method that uses minimal computer resources and which proves to be extremely efficient in achieving balance across 7 covariates of interest in a 2-arm CRT involving 32 primary care practices. Since this technique worked to achieve balance on 7 covariates, we can assume that it would have worked with fewer than 7 covariates.

This approach has limitations. While we did find an appropriately balanced randomization scheme, there is no guarantee that such a solution always exists. In such instances, we would recommend relaxing the balance criteria in some fashion or considering alternative approaches. For example, one might allow treatment groups to differ on a given factor, as long as a chi-square test for group differences remained above a specified threshold (e.g.  $p > 0.10$ ). Additionally, while this algorithm helps achieve marginal covariate balance, it does not ensure balance across other covariates within a given covariate's classification groups. For example, although the 2 residency programs are in different treatment groups, they are not balanced with respect to geographic region. By identifying only 116 acceptable randomization schemes, we may also have significantly reduced the statistical power for finding a significant treatment effect if we were to rely solely on permutation tests [21] in the analyses. However, the statistical power associated with hypothesis testing using general or generalized linear mixed models remains unaffected. Had many (e.g. thousands) more unique schemes satisfied the balance criteria, we might have been able to rely on a permutation test for the primary analysis, a statistical test often recommended for CRTs.

The strength of this technique is its computational efficiency in achieving marginal covariate balance across baseline covariates in a CRT. It may be particularly appealing when stratification on multiple factors is impractical. These methods may also be generalizable to studies involving 3 or more treatment group assignments, although more simulated randomization schemes may be required to achieve balance in such situations.

## Acknowledgements

This work was funded by a grant (no. 1 R01 CA112389-01A1) from the National Institute of Health, National Cancer Institute. We would also like to thank Valerie L Durkalski, PhD for her thoughtful review and critique of the manuscript prior to submission.

## References

1. Murray, DM. Design and Analysis of Group-Randomized Trials. New York: Oxford University Press; 1998.
2. Bland JM. Cluster randomised trials in the medical literature: two bibliometric surveys. BMC Med Res Methodol 2004;4:21. [PubMed: 15310402]

3. Raab GM, Butcher I. Balance in cluster randomized trials. *Stat Med* 2001;20:351–365. [PubMed: 11180306]
4. Institute of Medicine. *Crossing the Quality Chasm*. Washington, D.C: National Academy of Sciences; 2001.
5. Agency for Healthcare Research and Quality. *AHRQ Practice-Based Research Networks (PBRNs)*. Fact Sheet, June 2001 (revised May 2006). AHRQ Publication No. 01-P020. [Accessed 3-5-2008]. <http://www.ahrq.gov/research/pbrn/pbrnfact.htm>
6. Berner ES, Detmer DE, Simborg D. Will the wave finally break? A brief view of the adoption of electronic medical records in the United States. *J Am Med Inform Assoc* 2005;12:3–7. [PubMed: 15492029]
7. Moulton LH. Covariate-based constrained randomization of group-randomized trials. *Clin Trials* 2004;1:297–305. [PubMed: 16279255]
8. Tamura RN, Mills BJ, Lovelace JK. Constrained randomization in a therapeutic efficacy trial. *Biometrics* 1993;49:249–258. [PubMed: 8513107]
9. Berger VW. A review of methods for ensuring the comparability of comparison groups in randomized clinical trials. *Rev Recent Clin Trials* 2006;1:81–86. [PubMed: 18393784]
10. Chaudhary MA, Moulton LH. A SAS macro for constrained randomization of group-randomized designs. *Comput Methods Programs Biomed* 2006;83:205–210. [PubMed: 16870302]
11. Tiahrt KJ, Weeks DL. A method of constrained randomization for 2n factorials. *Technometrics* 1970;12:471–486.
12. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* 1975;31:103–115. [PubMed: 1100130]
13. Zucker DM, Lakatos E, Webber LS, et al. Statistical design of the Child and Adolescent Trial for Cardiovascular Health (CATCH): implications of cluster randomization. *Control Clin Trials* 1995;16:96–118. [PubMed: 7789139]
14. Piaggio G, Carroli G, Villar J, et al. Methodological considerations on the design and analysis of an equivalence stratified cluster randomization trial. *Stat Med* 2001;20:401–416. [PubMed: 11180310]
15. Gail MH, Byar DP, Pechacek TF, Corle DK. Aspects of statistical design for the Community Intervention Trial for Smoking Cessation (COMMIT). *Control Clin Trials* 1992;13:6–21. [PubMed: 1315664]
16. Hayes R, Mosha F, Nicoll A, et al. A community trial of the impact of improved sexually transmitted disease treatment on the HIV epidemic in rural Tanzania: 1 Design. *AIDS* 1995;9:919–926. [PubMed: 7576328]
17. Wight D, Raab GM, Henderson M, et al. Limits of teacher delivered sex education: interim behavioural outcomes from randomised trial. *BMJ* 2002;324:1430. [PubMed: 12065268]
18. Moore H, Summerbell CD, Greenwood DC, et al. Improving management of obesity in primary care: cluster randomised trial. *BMJ* 2003;327:1085. [PubMed: 14604931]
19. Nemeth LS, Feifer C, Stuart GW, Ornstein SM. Implementing change in primary care practices using electronic medical records: a conceptual framework. *Implement Sci* 2008;3:3. [PubMed: 18199330]
20. Feifer C, Ornstein SM. Strategies for increasing adherence to clinical guidelines and improving patient outcomes in small primary care practices. *Jt Comm J Qual Saf* 2004;30:432–441. [PubMed: 15357133]
21. Gail MH, Mark SD, Carroll RJ, Green SB, Pee D. On design considerations and randomization-based inference for community intervention trials. *Stat Med* 1996;15:1069–1092. [PubMed: 8804140]

**Table 1**  
Baseline comparison of practice characteristics among control and intervention practices.

	Control Practices (n=16)	Intervention Practices (n=16)
<b>Factors Used In Balancing Criteria</b>		
Geographic Region		
East: n (%)	4 (25.0%)	3 (18.8%)
South: n (%)	2 (12.5%)	3 (18.8%)
Midwest: n (%)	5 (31.3%)	5 (31.3%)
West: n (%)	5 (31.3%)	5 (31.3%)
State-Mandated CRC Screening Coverage		
No: n (%)	12 (75.0%)	12 (75.0%)
Yes: n (%)	4 (25.0%)	4 (25.0%)
Baseline CRC Performance Tertile		
Low: n (%)	5 (31.3%)	6 (37.5%)
Middle: n (%)	5 (31.3%)	5 (31.3%)
High: n (%)	6 (37.5%)	5 (31.3%)
Past network experience		
No: n (%)	12 (75.0%)	12 (75.0%)
Yes: n (%)	4 (25.0%)	4 (25.0%)
Specialty		
Internal Medicine: n (%)	3 (18.8%)	2 (12.5%)
Family Medicine: n (%)	13 (81.3%)	14 (87.5%)
Number of providers		
1 or 2: n (%)	6 (37.5%)	5 (31.3%)
3 or 4: n (%)	5 (31.3%)	5 (31.3%)
5 or more: n (%)	5 (31.3%)	6 (37.5%)
Residency program		
No: n (%)	15 (93.8%)	15 (93.8%)
Yes: n (%)	1 (6.3%)	1 (6.3%)
<b>Other Factors</b>		
Number of patients 50 years and older per practice		
Median (Interquartile Range)	982 (743 – 2460)	1260 (752 – 2388)
Percent of patients 50 years and older up-to- date with CRC screening		
Median (Interquartile Range)	51.6% (44.6% – 56.0%)	50.8% (44.8% – 57.4%)
Average age of patients 50 years and older in each practice		
Median (Interquartile Range)	64.5 (62.7 – 67.2)	64.8 (63.6 – 67.0)
Percent of patients 50 years and older who are male		
Median (Interquartile Range)	43.0% (38.9% to 47.3%)	40.9% (38.9% to 49.1%)