SPECIAL COMMUNICATION

Pragmatic Cluster Randomized Trials Using Covariate Constrained Randomization: A Method for Practice-based Research Networks (PBRNs)

L. Miriam Dickinson, PhD, Brenda Beaty, MSPH, Chet Fox, MD, Wilson Pace, MD, W. Perry Dickinson, MD, Caroline Emsermann, MS, and Allison Kempe, MD, MPH

Background: **Cluster randomized trials (CRTs) are useful in practice-based research network translational research. However, simple or stratified randomization often yields study groups that differ on key baseline variables when the number of clusters is small. Unbalanced study arms constitute a potentially serious methodological problem for CRTs.**

Methods: **Covariate constrained randomization with data on relevant variables before randomization was used to achieve balanced study arms in 2 pragmatic CRTs. In study 1, 16 counties in Colorado were randomized to practice-based or population-based reminder recall for vaccinating children ages 19 to 35 months. In study 2, 18 primary care practices were randomized to computer decision support plus practice facilitation versus computer decision support alone to improve care for patients with stage 3 and 4 chronic kidney disease. For each study, a set of optimal randomizations, which minimized differences of key variables between study arms, was identified from the set of all possible randomizations.**

Results: **Differences between study arms were smaller in the optimal versus remaining randomizations. Even for the randomization in the optimal set with the largest difference between groups, study arms did not differ significantly on any variable for either study ($P > .05$).**

Conclusions: **Covariate constrained randomization, which restricts the full randomization set to a subset in which differences between study arms are minimized, is a useful tool for achieving balanced study arms in CRTs. Because of the increasing recognition of the risk of imbalance in CRTs and implications for interpreting study findings, procedures of this type should be considered in designing practice-based or community-based trials. (J Am Board Fam Med 2015;28:663–672.)**

Keywords: **Cluster Randomized Trials, Covariate-Based Constrained Randomization, Practice-based Research**

Practice-based research networks (PBRNs) are ideal settings for pragmatic clinical trials and implementation and dissemination research, supporting the goal of moving evidence-based interventions for common health problems into practice.[1,2] A major advantage of pragmatic trials is that they are done in real-world settings and populations; thus findings are more readily generalizable. Key characteristics of pragmatic trials include (1) comparison of clinically relevant alternative interventions; (2) diverse populations of study participants,

This article was externally peer reviewed.

Submitted 1 January 2015; revised 28 March 2015; accepted 13 April 2015.

From the Department of Family Medicine, University of Colorado, Denver (LMD, WP, WPD, CE); the Adult and Child Center for Outcomes Research and Delivery Science (ACCORDS), University of Colorado, Denver (LMD, BB, AK); the American Academy of Family Physicians National Research Network, Leawood, KS (LMD, WP); the DARTNet Institute, Aurora, CO (CH, WP); Department of Family Medicine, the State University of New York, Buffalo (CF); the Department of Pediatrics, University of Colorado, Denver (AK); and the Children's Hospital Colorado, Aurora, CO (AK).

Funding: Funding for this work was provided by the Agency for Healthcare Research and Quality (grant no. P01HS021138) and the National Institute of Diabetes and Digestive and Kidney Diseases (grant no. R01 DK090407).

Conflict of interest: none declared.

Prior Presentation: Some of these data were presented at the North American Primary Care Research Group meeting, New Orleans, LA (December 2012); and the World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (WONCA) meeting, Prague, Czech Republic (June 2013).

Corresponding author: L. Miriam Dickinson, PhD, Department of Family Medicine, University of Colorado School of Medicine, 12631 E. 17th Ave., Mail Stop F496, Aurora, CO 80045-0508 (E-mail: miriam.dickinson@ucdenver.edu).

J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.

similar to individuals who are affected by the condition(s) being studied; (3) heterogeneous practice settings similar to those where the condition is generally treated; and (4) collection of data on a broad range of health outcomes.[3,4]

Cluster randomized trials (CRTs) are often the most feasible study design for pragmatic trials. Interventions that are designed to be at the level of a medical practice or community necessitate a unit of randomization other than the individual. Simple randomization, generally executed using a single sequence of random assignment via a random number generator or table, is expected to produce comparable study arms when the number of units to be randomized is sufficiently large (>30 per arm). When there are relatively few clusters with diverse settings and populations, however, simple randomization can result in study arms that differ substantially on key process and clinical variables or potential confounders, such as sociodemographic or contextual features.[5–9] Imbalance in a trial weakens the case for causal inference, the key strength of a randomized controlled trial, because observed differences between study arms following an intervention may not be attributable to the intervention, but rather may be the result of differences in the underlying study populations.[5] In addition, imbalance can result in differences between crude and adjusted estimates of treatment effects, thus hindering the interpretability and face validity of the findings.[5,10–12] This problem has resulted in questions regarding the validity and/or generalizability of results from CRTs.

Imbalance in cluster or individual characteristics occurs fairly often[10–15] in CRTs when simple randomization is used. Restricted randomization methods may improve the chances of achieving balanced study arms. Stratification (block randomization) and matching have most commonly been used, but these are feasible for only a limited number of categorical variables[16] and may be ineffective in achieving balance.[5,11] In addition, stratifying or matching on combinations of multiple variables is often not even possible. Covariate-constrained randomization,[17–23] a more complex but particularly promising approach, is feasible when the number of variables is larger or includes continuous variables. This procedure, which reduces the set of all possible randomizations to a subset of randomizations in which differences between study arms have been minimized, has been shown to achieve excellent balance in baseline characteristics and improved power in simulation studies.[24,25] Despite this, it is seldom used in practice.[5] For this approach, investigators need a sufficient number of clusters (eg, practices, communities) to avoid overconstraint.[20] A minimum of 8 clusters is recommended.[5,22] In addition, baseline data on cluster and individual characteristics must be available to carry out the randomization procedure. In a recent methodological review of allocation techniques for CRTs, Ivers et al[5] commented that "covariate-constrained randomization can offer investigators the chance to remove the risk of baseline imbalance with minimal risk for bias." They note that the infrequent use of such procedures represents "a gap in methodological best practices."[5]

The purpose of this article is to illustrate the use of covariate-constrained randomization as a method for producing a subset of randomizations with comparable study arms in 2 pragmatic CRTs. Study 1 used a simple covariate constrained randomization approach for rural and urban counties in Colorado, using data from the 2010 Census and a state-based immunization registry. Study 2 incorporated stratification variables into the randomization procedure for primary care practices across several regions of the United States. We hypothesized that use of covariate-constrained randomization would be feasible and would improve the overall chances of achieving balanced study arms. This would also alleviate the risk of extreme imbalance in study arms in terms of key cluster and individual patient characteristics and improve internal and external validity of the study.[26]

Methods

Cluster Randomized Trials

Study 1 is a CRT of 2 reminder-recall (R/R) approaches (population vs practice-based) for increasing up-to-date immunization rates in 19- to 35-month-old children in 16 counties in Colorado, as part of the Agency for Health care Research and Quality–funded Center for Research in Implementation Science and Prevention (grant P01HS021138; A. Kempe, primary investigator). Identification of the patient cohort, practice affiliation, and outcomes were ascertained through the Colorado Immunization Information System (CIIS). All practices from study counties that delivered immunizations to 19- to 35-month-old children and were enrolled in CIIS were

J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.

included.[27] The unit of randomization for this study was the county. The population-based intervention used collaborations between primary care physicians, CIIS, and health department leaders to develop a centralized R/R notification (telephone and mail) for all parents whose 19- to 35-month-old children were not up to date on immunizations. In the practice-based arm, practices were invited to attend a webinar training on R/R using CIIS and were offered financial reimbursement for mailing or autodialer costs for R/R notifications.

Study 2 is a National Institute of Diabetes, Digestive and Kidney Diseases–funded study (grant R01 DK090407, C. Fox, primary investigator) of computer decision support (CDS) plus practice facilitation versus CDS alone to improve care and outcomes for patients with stage 3 and 4 chronic kidney disease (CKD) in primary care practices recruited from the DARTNet Collaborative set of networks.[28] In addition to the CDS tool that all practices received, practices in the facilitation arm received assistance with site coordination, physician champion's needs, audit and feedback, team approach to care, and education.[28] Randomization of the 18 practices from the first wave is described here. Both studies are registered with clinicaltrials.gov (identifiers NCT01557621 and NCT01767883).

Covariate-Constrained Randomization: General Procedures

To use covariate-constrained randomization, baseline data on clusters must be available. Investigators must identify a set of variables that (1) may be associated directly with the study outcome(s), or (2) may be potential confounders, or (3) could affect implementation of the intervention in practice or community settings. These can be measured directly at the cluster level (eg, rural location, practice size or type), or they may be derived from aggregated patient- or population-level data (eg, mean age, race/ethnicity, mean or percentage at goal on clinical measures). Once cluster-level data are obtained, the randomization process can begin.

The first step for a 2-arm trial is to generate all possible randomizations of clusters into 2 study arms. For convenience we refer to them here as study groups 1 and 2. In the applications we describe here, the IML Procedure in SAS (SAS Institute Inc., Cary, NC) was used (see Online Appendix, step 1).[23] Each cluster is assigned a numeric id (1, 2, 3, . . . n).

If there are stratification variables, they can be incorporated into the procedure at this time (study 2), or investigators may choose to perform separate randomizations for different strata (study 1) and then combine them. Decisions about stratification requirements should be made in advance and usually involve requiring equal (or nearly equal) distributions of certain cluster characteristics in each study arm. This can be achieved by retaining only randomizations that have the required combination of clusters in each arm. For example, investigators might specify in advance that they want an equal number of rural practices in each study arm. Using the assigned cluster ID to create an indicator variable, the number of rural practices in each study group can be determined for each randomization, and those that do not have the specified number of rural practices (eg, 4 per study group) per arm are removed from the set of possible randomizations (see Online Appendix, step 2).

To allow each measure to contribute approximately equally to the balancing process, cluster-level variables should be standardized (Online Appendix, step 3). Here we use a simple z score,

$$z_i = \frac{x_i - \bar{x}}{S}$$

where x_i is the practice-level variable and s is the standard deviation across practices. The study 1 variables (cluster-level means, counts, percentages) that were converted to z scores are shown in the Online Appendix. The standardized cluster data set is replicated and merged with the randomization data set (Online Appendix, steps 4 and 5). The cluster ID is then used to determine which clusters are in study group 1 for each randomization generated by PROC IML; clusters that are not chosen for study group 1 are automatically assigned to study group 2 (Online Appendix, step 6).

Standardized variables then are used to compute a balance criterion for each randomization, defined as the sum of squared differences between group means on all variables included in the balancing procedure (Online Appendix, steps 7 to 9). The balance criterion is actually a measure of the overall imbalance, or distance, between the allocated study groups; larger values indicate greater imbalance,

J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.

Table 1. Study 1 County-Level Baseline Variables for Randomization

Variable	County-Level (Rural and Urban) Variables for Randomization (n = 16)		Differences in Raw Variables Between Study Groups*	
	Mean (SD)	Minimum, Maximum	Optimal Randomizations (n = 14)	Remaining Randomizations (n = 126)
In CIIS (%)[†]	87.2 (7.7)	70, 100	<1 (1)	<1 (1)
Children aged 19–35 months, by county (n)	4,197 (4,432)	234, 12,354	223 (613)	1,264 (6,325)[‡]
Up to date at baseline (%)	40.8 (8.3)	27.0, 54.0	2.1 (5.0)	4.9 (15.0)[‡]
Hispanic ethnicity (%)[§]	22.3 (12.9)	6.0, 44.0	5.6 (11.3)	7.9 (23.3)
African American race[‖]	2.9 (2.7)	0, 10.0	<1 (1.0)	1.4 (4.5)
Average income ($)[¶]	53,481 (15,793)	29,738, 93,819	3,659 (9,702)	9,731 (27,131)[‡]
Pediatric–to–Family Medicine practices ratio	0.28 (0.25)	0, 1.0	0.20 (0.40)	0.15 (0.40)
CHCs (n)	4.4 (3.5)	0, 11	1.3 (2.8)	1.6 (4.8)

Data are based on baseline information on the counties from the Colorado Department of Public Health and Environment, the Colorado Immunization Information System (CIIS), and 2010 Census data.
*Data are means (maximums).
[†]Percentage of children between 0 and 4 years who had ≥2 immunization records in CIIS.
[‡]$P < .01$, optimal vs remaining randomizations.
[§]Population by ethnicity (Hispanic/Latino or not).
[‖]Population by race (black or other).
[¶]Population by income.
CHC, community health center; SD, standard deviation.

whereas smaller values indicate greater balance. Weights (w_i) can be applied to each variable, if desired.

$$B = (w_1(z_{11} - z_{21})^2 + w_2(z_{12} - z_{22})^2 + \ldots),$$

where z_{11} is the mean of group 1 units on standardized variable 1 and z_{21} is the mean of group 2 units on standardized variable 1, and so on.

A cut point is established for the maximum allowable difference between study groups, based on the maximum allowable value for B, to define a set of "acceptable randomizations," which we call the "optimal set," in which the differences between study groups on covariates are minimized (Online Appendix, step 10). Although there is no set standard at this time, the criterion used here was to define approximately the best 10% as the optimal set (the 10th percentile of the distribution of B), thus minimizing the distance between designated study groups.

Finally, a single randomization is randomly selected from the set of optimal randomizations to allocate clusters to study arms in the trial (Online Appendix, step 11).

Variables Used to Calculate the Balance Criterion

For study 1, county-level variables were used in the randomization procedure, and rural and urban counties were randomized separately. County-level data obtained from the US Census and the Colorado Immunization Information System are shown in Table 1. Variables were weighted equally (ie, $w_i = 1$) and standardized (z scores) before computing the balance criterion.

For study 2, practice-level variables determined from direct report or created by identifying potentially eligible patients with stage 3 or 4 CKD before baseline using electronic health record (EHR) data, and aggregated to the level of the practice to obtain practice-level means and rates, are shown in Table 2. Again, practice-level variables were weighted equally ($w_i = 1$) and standardized (z scores) before computing the balance criterion. Practices were located in 3 geographic regions, and there were 8 practices that were members of 3 organizations. These were incorporated into the randomization as stratification variables.

For both studies, the "acceptable set" was defined as approximately the best 10% of randomizations, based on the values of the balance criterion B. Randomizations with the lowest values for B were chosen for the optimal set.

J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.

Table 2. Study 2 Practice-Level Baseline Variables for Randomization

Practice-Level and Clinical Variables Obtained from EHR Data	Practice-Level Variables for Randomization (n = 18 Practices)		Differences Between Study Groups on Raw Variables*	
	Mean (SD)	Minimum, Maximum	Optimal Randomizations (n = 172)	Remaining Randomizations (n = 1556)
FTE clinicians (n)	3.8 (3.1)	1, 12	1.0 (2.6)	1.3 (3.2)†
African American race (%)	2.9 (3.7)	<1, 15.0	1.2 (3.1)	1.4 (3.1)‡
Hispanic ethnicity (%)	16.6 (18.5)	<1, 68.0	3.6 (10.7)	3.9 (10.7)
Medicaid/Uninsured (%)	13.8 (8.6)	0, 30.0	1.8 (3.3)	3.0 (7.5)†
Have type 2 diabetes (%)	34.9 (18.4)	13.0, 100	8.4 (15.3)	9.1 (16.8)†
HbA_{1c} >9 (%)	9.1 (7.5)	0, 25	1.5 (4.8)	2.4 (8.1)†
Stage 4 CKD (%)	6.7 (4.5)	0, 15.6	<1 (2.6)	1.9 (5.6)†
BP >130/80§ (%)	58.1 (14.2)	35.4, 1.0	5.7 (14.4)	5.8 (18.0)†
BP>140/90 mmHg‖ (%)	31.2 (11.6)	0, 51.9	2.6 (6.2)	4.5 (14.8)†
Mean HbA_{1c}¶	7.02 (0.38)	6.63, 7.88	0.05 (0.15)	0.12 (0.33)†
Mean eGFR	49.4 (3.2)	45.2, 59.4	0.7 (2.2)	1.3 (4.0)†
Mean systolic BP	132.1 (3.8)	124.7, 138.6	0.8 (2.5)	1.6 (5.1)†

*Data are means (maximums).
†$P < .01$, optimal vs. remaining randomizations.
‡$P < .05$, optimal vs. remaining randomizations.
§Patients have either systolic blood pressure (BP) >130 or diastolic BP >80, or both.
‖Patients have either systolic BP >140 or diastolic BP >90, or both.
¶Includes diabetic patients only.
CKD, chronic kidney disease; eGFR, estimate glomerular filtration rate; FTE, full-time equivalent; SD, standard deviation.

Comparison of Optimal Set with Remaining Randomizations

For each possible randomization, clusters were allocated to a treatment group (group 1 or group 2). Next, the absolute value of the difference between the allocated treatment groups on the original county- and practice-level raw variables was computed for each randomization ($D = abs[x_{11} - x_{21}]$). These differences between the optimal set and the remaining randomizations were compared using Wilcoxon rank-sum tests to determine whether the procedure improved balance on the original variables.

Results

Study 1

County-level data on the 16 rural and urban counties in Colorado shown in Table 1 indicate considerable heterogeneity. Within each stratum (rural, urban), SAS Proc IML generated 70 possible combinations of 8 counties into 2 study groups (This includes 70 urban and 70 rural, for a total of 140.). Next, county-level variables were standardized by computing z scores for each measure. For each randomization, the balance criterion was calculated as described above. The best 10% of randomizations (ie, lowest values for B) for rural and urban counties were designated as the optimal set. Figure 1 shows the distribution of the balance criterion across randomizations for the optimal set, the remaining randomizations, and the full randomization set (total area under the curve). From this figure, it can be seen that, although it is possible to obtain balanced study arms using simple or stratified randomization (ie, full randomization set), there is a considerable chance of obtaining highly imbalanced (large values for B) study arms. Essentially, covariate-constrained randomization increases the probability of obtaining a balanced randomization by limiting the possibilities to a subset with reasonable balance.

The comparison of the optimal set with the remaining randomizations is shown in Table 1. The magnitude of differences between groups on the raw variables for the optimal set versus remaining randomizations indicate that the average distance between study groups, as well as the maximum difference between study groups, was generally smaller for the optimal set compared with the remaining randomizations (significantly smaller for the number of children ages 19 to 35 months, percentage up to date at baseline, and average income).

J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.

Figure 1. Distribution of balance criterion in study 1. The full set of all possible randomizations is represented by the total area under the curve.

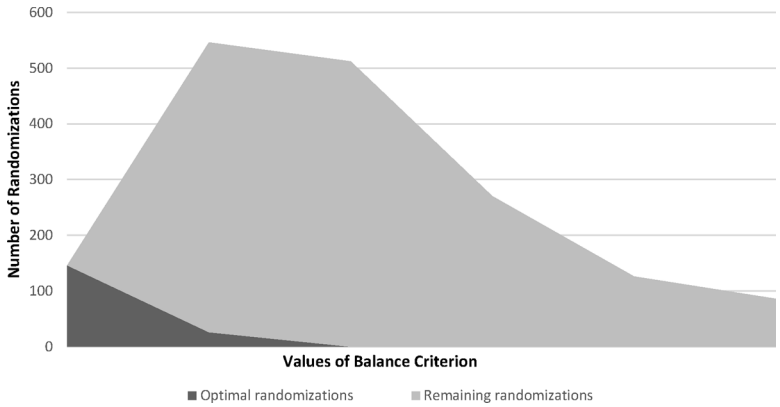


Table 3. Study 1: County-Level Baseline Variables by Group for Worst Randomization from Optimal Set

	County-Level Variables, Mean (SD)	
Variable	Group 1	Group 2
In CIIS (%)*	87.8 (10.1)	86.6 (5.0)
Children aged 19–35 months (n)	4,275 (4628)	4,118 (4546)
Up to date at baseline (%)	40.1 (8.8)	41.5 (8.3)
Hispanic ethnicity (%)	23.8 (14.8)	20.9 (11.6)
African American race (%)	2.5 (2.4)	3.3 (3.1)
Average income ($)	56,264 (18,004)	50,699 (13,877)
Pediatric–to–Family Medicine ratio	0.33 (0.33)	0.23 (0.15)
CHCs (n)	4.8 (4.5)	4.0 (2.4)

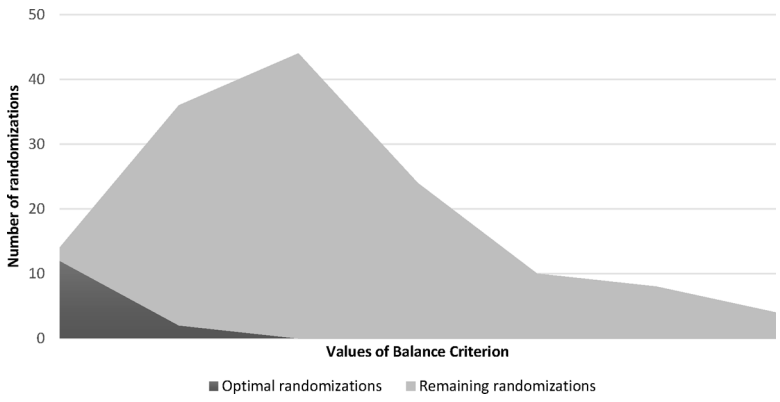
*Colorado Immunization Information System.
CHC, community health center; SD, standard deviation.

Table 3 shows group means (standard deviations) on raw variables for the worst randomization (largest B) from the optimal set. There were no significant differences on any variable between study arms.

Study 2

Practice-level data for the 18 practices in the CKD study are shown in Table 2, again demonstrating considerable heterogeneity. This study included stratification variables in the procedure, rather than obtaining randomizations separately for each stratum and then combining. Initially, there were 48,620 possible combinations of 18 practices into 2 groups. Stratification variables included geographic location and practice organization. We determined the number of practices that were needed in each group for the 3 geographic locations and considered only randomizations with the correct alloca-

Figure 2. Distribution of balance criterion. The set of all possible randomizations (after applying stratification criteria) is represented by the total area under the curve.



J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.

Table 4. Study 1: Practice-Level Baseline Variables by Group for Worst Randomization from Optimal Set

Variable	Practice-Level Variables, Mean (SD)	
	Group 1	Group 2
FTE clinicians (n)	3.9 (2.9)	3.6 (3.6)
African American race (%)	2.2 (1.2)	3.7 (5.2)
Hispanic ethnicity (%)	13.0 (3.6)	20.1 (2.1)
Medicaid/Uninsured (%)	13.1 (8.2)	14.4 (9.4)
Diabetic (%)	30.4 (11.6)	39.3 (23.3)
HbA_{1c} >9 (%)	10.3 (8.9)	8.0 (6.2)
Stage 4 CKD (%)	7.3 (3.8)	6.1 (5.3)
BP >130/80 (%)	55.4 (6.7)	60.1 (19.1)
BP >140/90 (%)	33.3 (6.5)	29.2 (15.4)
Mean HbA_{1c}	7.0 (0.4)	7.0 (0.4)
Mean eGFR	48.4 (2.4)	50.4 (3.6)
Mean systolic BP	132.3 (3.0)	131.9 (4.6)

BP, blood pressure; CKD, chronic kidney disease; eGFR, estimate glomerular filtration rate; FTE, full-time equivalent; SD, standard deviation.

tion. A similar approach was taken for the 3 practice organizations. There were a total of 1728 randomizations that met stratification criteria. Next, practice-level variables were standardized (z scores). The balance criterion was computed for each randomization and, after examining the distribution of the balance criterion, the optimal set was defined as approximately the best 10%. Table 2 shows the average and maximum differences on each variable for optimal and remaining randomizations between study groups. Differences were significantly smaller in optimal randomizations compared with the remaining randomizations for most variables. Figure 2 shows the distribution of the balance criterion for the optimal and remaining sets. As before, comparison of the optimal set with the full randomization set (total area under the curve) shows that the probability of obtaining a well-balanced randomization is greater using the covariate-constrained randomization approach. Finally, study arms assigned using the worst randomization (largest B) in the optimal set were compared based on raw variables (Table 4). Again, there were no significant differences on any variable between study arms.

Discussion

Achieving balanced study arms should be an important priority when designing and implementing pragmatic CRTs. Stratification and matching, the most commonly used strategies, are often insufficient and result in nonequivalent study arms. Imbalance in study arms can compromise an investigator's ability to draw conclusions about intervention effectiveness because of underlying differences in the populations.[1] In addition, interpretation of observed treatment effects is more difficult when study arms are unbalanced on baseline covariates because crude and adjusted estimates may differ considerably.[1,6–8] In this article we have described 2 applications of covariate-constrained randomization, a procedure for achieving balance across study arms in CRTs.

Using covariate-constrained randomization, we identified a set of optimal randomizations from which the final study allocation was randomly selected. For the 2 studies described here, cluster-level data were obtained before randomization from a variety of sources, including EHRs, direct practice reporting, a state immunization registry, and US Census data. We demonstrated that differences on most cluster-level variables between study arms were significantly less for the optimal versus remaining randomizations. Perhaps more important, the chance of obtaining a well-balanced study arm allocation was much greater using this procedure.

An element critical to the success of this approach is thoughtful selection of variables to be used in calculating the balance criterion. Both information from previous studies and clinical insight should be applied when selecting the variables. The procedure can accommodate dichotomous or categorical variables as well as continuous variables, which allows clinically logical choices concerning relevant characteristics. Covariate-constrained randomization allows the inclusion of a larger number of variables than is possible with stratification, where decisions regarding the few, most important variables can be difficult and sometimes arbitrary. However, the ability to include many variables should be weighed against the fact that including potentially irrelevant variables will, in effect, lessen the influence of the remaining, more important variables. Thus it may be more prudent to include fewer, carefully selected variables when there is strong evidence regarding which patient, practice, and contextual characteristics are likely to be most important. Regardless of the number of variables included, covariate-constrained randomization fa-

J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.

cilitates balancing those variables across the study arms to a level generally not possible with stratification.[30,31]

The 2 CRTs described here illustrate the variable selection process. In study 1, rural and urban differences were considered to be so extensive and multifaceted that the decision was made to randomize in separate strata. This was based on data demonstrating major differences in socioeconomic profiles, insurance coverage, provider specialties, the way health care is delivered in rural versus urban areas, and where immunizations are delivered (more likely to be in public settings or via public nursing entities in rural settings). Each of these issues could affect the potential for increasing immunization rates, independent of the specific intervention being studied. Investigators included the ratio of pediatric to family medicine practices because of known differences in the uptake of immunization delivery interventions by specialty. Racial/ethnic differences were important to include because of known differences in health-seeking behavior, access to care, and insurance coverage. Baseline up-to-date rates were important because of a known ceiling effect in the ability to increase immunization rates. Although not all variables of potential importance were available by county, the constellation of variables available was considered to adequately account for major sources of potential confounding.

In study 2, numerous clinical variables were available from EHR data. Though inclusion criteria restricted the patient population to patients with stage 3 or 4 CKD, there was heterogeneity among practices regarding the extent of renal function decline, so average estimated glomerular filtration rates and percentage with stage 4 CKD were included. Since many patients had diabetes and hypertension, average HbA_{1c} and systolic blood pressure (BP) were included. Data on some common quality indicators (ie, percentage with HbA_{1c} >9, percentage with BP >140/90 mmHg, percentage with BP >130/80 mmHg) were thought to be important because of the potential impact on improvement in outcomes, and as indicators of practice success in quality improvement efforts. Practice characteristics such as organizational structure and the sociodemographics of populations served were also considered to be important because they may affect a practice's ability to implement interventions or achieve desired effectiveness.

Several limitations of the covariate-constrained randomization procedure should be noted. Prerandomization data on key characteristics are required and may often be difficult to obtain. The utility of this approach is only as good as the data that are used in the procedure. There has been little systematic research regarding several decision points in the process, including how many variables should be used, how weighting can be used to emphasize the most important variables, and how the procedure can be extended to include more than 2 study arms. Finally, when the number of clusters is large, the number of possible randomizations can be prohibitive. For example, randomization of 30 practices into 2 groups of 15 yields 155,117,520 possible randomizations. For larger sample sizes we have used stratification or blocking to obtain smaller groups for randomization before using the procedures described here. Other investigators have encountered this problem as well, and promising alternative approaches have been developed.[32]

Applying methods for attaining balanced study arms enhances the ability to learn from pragmatic CRTs. Since pragmatic trials are conducted in real-world settings with patients with multiple and varied comorbidities, the results have real-world applicability; yet there are challenges in designing and executing CRTs that must be addressed. Covariate-constrained randomization is a feasible approach to achieving balanced study arms. While this procedure requires additional time and effort, it can help researchers get the most out of these expensive and time-consuming endeavors and help foster the production of high-quality, generalizable knowledge. Furthermore, minimizing differences between study arms generally has the effect of increasing variability within study arms. For larger trials, this may improve the researcher's ability to examine factors that affect intervention implementation as well as differential effectiveness.[32] Further research is needed to determine optimal conditions for applying this procedure and to explore benefits and potential drawbacks.

Conclusion

We recommend the use of covariate-constrained randomization approaches to improve balance in

J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.

pragmatic CRTs when pre-randomization data on clusters are available and the number of clusters to be randomized is adequate. Because of widespread recognition of the risk of imbalance in CRTs and the difficulties this problem poses for interpreting study findings, procedures of this type are increasingly becoming a necessary state of the art in the design of practice-based or community-based trials.

References

1. Heintzman J, Gold R, Krist A, Crosson J, Likumahuwa S, DeVoe JE. Practice-based research networks (PBRNs) are promising laboratories for conducting dissemination and implementation research. J Am Board Fam Med 2014;27:759–62.
2. Glasgow RE, Vinson C, Chambers D, Khoury MJ, Kaplan RM, Hunter C. National Institutes of Health approaches to dissemination and implementation science: current and future directions. Am J Public Health 2012;102:1274–81.
3. Cocannon TW, Guise JM, Dolor RJ, et al. A national strategy to develop pragmatic clinical trials infrastructure. Clin Transl Sci 2014;7:164–71.
4. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA 2003;290:1624–32.
5. Ivers NM, Halperin IJ, Barnsley J, et al. Allocation techniques for balance at baseline in cluster randomized trials: a methodological review. Trials. 2012;13:120.
6. Ivers NM, Taljaard M, Dixon S, et al. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000–8. BMJ 2011;343:d5886.
7. Eldridge SM, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomized trials: systematic review of recent trials. BMJ 2008;336:876–80.
8. Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Uloumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. Clin Trials 2004;1:80–90.
9. Donner A, Klar N. Design and analysis of cluster randomization trials in health research. New York, NY: Oxford University Press; 2000.
10. Alexander FE, Roberts MM, Lutz W, Hepburn W. Randomization by cluster and the problem of social class bias. J Epidemiol Community Health 1989;43:29–36.
11. Shah S, Peat JK, Mazurski EJ, et al. Effect of peer led programme for asthma education in adolescents: cluster randomized controlled trial. BMJ 2001;322:583–5.
12. Carman WF, Elder AG, Wallace LA, et al. Effects of influenza vaccination of health-care workers on mortality of elderly people in long-term care: a randomized controlled trial. Lancet 2000;355:93–7.
13. Solomon DH, Polinski JM, Stedman M, et al. Improving care of patients at-risk for osteoporosis: a randomized controlled trial. J Gen Intern Med 2007;22:362–7.
14. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomized trials: review of recent trials published in three general medical journals. BMJ 2003;327:785–9.
15. Chapman S, Cornwall J, Righetti J, Sung L. Preventing dog bites in children: randomised controlled trial of an educational intervention. BMJ 2000;320:1512–3.
16. Zhao W, Hill MD, Palesch Y. Minimal sufficient balance–a new strategy to balance baseline covariates and preserve randomness of treatment allocation. Stat Methods Med Res. 2012 January 26 [Epub ahead of print]
17. Raab GM, Butcher I. Balance in cluster randomized trials. Stat Med 2001;20:351–65.
18. Kraschnewski JL, Keyserling TC, Bangdiwala SI, et al. Optimized probability sampling of study sites to improve generalizability in a multisite intervention trial. Prev Chronic Dis 2010;7:A10.
19. Glynn RJ, Brookhart A, Stedman M, Avorn J, Solomon DH. Design of cluster-randomized trials of quality improvement interventions aimed at medical care providers. Med Care 2007;45(10 Suppl 2):S38–43.
20. Moulton LH. Covariate-based constrained randomization of group-randomized trials. Clin Trials 2004;1:297–305.
21. Chaudhary MA, Moulton LH. A SAS macro for constrained randomization of group-randomized designs. Comput Methods Programs Biomed 2006;83:205–10.
22. Carter BR, Hood K. Balance algorithm for cluster randomized trials. BMC Med Res Methodol 2008;8:65.
23. Beaty B, Dickinson LM. Adjusting clustering: minimize your suffering! Improving balance in baseline covariates in cluster randomized trials. Available from: http://support.sas.com/resources/papers/proceedings14/1570-2014.pdf. Accessed July 16, 2015.
24. Perry M, Faes M, Reelick MF, Olde Rikkert MG, Borm GF. Studywise minimization: A treatment allocation method that improves balance among treatment groups and makes allocation unpredictable. J Clin Epidemiol 2010;62:1118–22.
25. Xiao L, Lavori PW, Wilson SR, Ma J. Comparison of dynamic block randomization and minimization in randomized trials: a simulation study. Clin Trials 2011;8:59–69.
26. Giraudeau B, Ravaud P. Preventing bias in cluster randomised trials. PLoS Med 2009;6:e1000065.

J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.

27. Kempe A, Saville AW, Dickinson LM, et al. Collaborative centralized reminder/recall to increase immunization rates for populations of young children: a comparative effectiveness trial. JAMA Pediatr. In press.
28. Fox CH, Vest BM, Kahn LS, et al. Improving evidence-based primary care for chronic kidney disease: Study protocol for a cluster randomized control trial for translating evidence into practice (TRANSLATE CKD). Implement Sci 2013;8:88.
29. Dickinson WP, Dickinson LM, Nutting PA, et al. Practice facilitation to improve diabetes care in primary care: a report from the EPIC randomized clinical trial. Ann Fam Med 2014;12:8–16.
30. Dickinson LM, Dickinson WP, Nutting PA, et al. Practice context affects efforts to improve diabetes care for primary care patients: a pragmatic cluster randomized trial. J Gen Intern Med 2015;30:476–82.
31. Nietert PJ, Jenkins RG, Nemeth LS, Ornstein SM. An application of a modified constrained randomization process to a practice-based cluster randomized trial to improve colorectal cancer screening. Contemp Clin Trials 2009;30:129–32.

J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.

APPENDIX

Sample SAS Code for Each Step

1. Generate All Possible Randomizations

SAS code for the IML procedure (this can also be done using PROC PLAN). This code creates a data set that contains 9 variables (col1 to col9), each with values ranging from 1 to 18. These values are the practice ID numbers that are arbitrarily assigned to the 18 practices in this randomization set. The 9 selected for each randomization are group 1. The remaining practices ultimately are assigned to group 2.

```
proc iml;
n = 18;
k = 9;
c = allcomb(n,k);
create out from c;
append from c;
quit;
```

2. Balance on Stratification Variables (If Applicable)

```
*In this example "rural" is a stratification variable: practices 1, 3, 5, and 6 are rural;
*subset on stratification first to reduce number of acceptable randomizations; *need 2 rural practices per group;
data sasfiles.rand18;
set out;
rural = 0;
if col1 = 1 or col2 = 1 or col3 = 1 or col4 = 1 or col5 = 1 or col6 = 1 or col7 = 1 or col8 = 1 or col9 = 1 then rural = rural+1;
if col1 = 3 or col2 = 3 or col3 = 3 or col4 = 3 or col5 = 3 or col6 = 3 or col7 = 3 or col8 = 3 or col9 = 3 then rural = rural+1;
if col1 = 5 or col2 = 5 or col3 = 5 or col4 = 5 or col5 = 5 or col6 = 5 or col7 = 5 or col8 = 5 or col9 = 5 then rural = rural+1;
if col1 = 6 or col2 = 6 or col3 = 6 or col4 = 6 or col5 = 6 or col6 = 6 or col7 = 6 or col8 = 6 or col9 = 6 then rural = rural+1;
run;
*Here we add an organizational variable (orgA); practices 8 to 11 belong to the same organization, and we want 2 practices from orgA in each group;
data sasfiles.rand18a;
set sasfiles.rand18;
OrgA = 0;
if col1> = 8 and col1< = 11 then OrgA = OrgA+1;if col2> = 8 and col2< = 11 then OrgA = OrgA+1;if col3> = 8 and col3< = 11 then OrgA = OrgA+1;
if col4> = 8 and col4< = 11 then OrgA = OrgA+1;if col5> = 8 and col5< = 11 then OrgA = OrgA+1;if col6> = 8 and col6< = 11 then OrgA = OrgA+1;
if col7> = 8 and col7< = 11 then OrgA = OrgA+1;if col8> = 8 and col8< = 11 then OrgA = OrgA+1;if col9> = 8 and col9< = 11 then OrgA = OrgA+1;
run;
proc freq data = sasfiles.rand18a;
tables rural orgA;
run;
*Only keep randomizations with exactly 2 rural and 2 orgA practices;
data sasfiles.rand18a;
set sasfiles.rand18a;
if rural = 2 and orgA = 2 then output;
run;
*Add a randomization number to each randomization;
data sasfiles.rand18a;
set sasfiles.rand18a;
rand+1;
run; *Note how many randomizations there are after applying stratification criteria (N) for step 4;
```

3. Standardize Cluster Level Variables to Create z Scores Using PROC Standard

The example below uses only 4 variables (for brevity); original variables are replaced with their *z* scores.

```
Proc standard data = clusterdata mean = 0 std = 1 out = zscores;
var var1—var4;
Run;
```

4. Create n Copies of Standardized Cluster Data

```
data r;
set sasfiles.clusterdata;
run;
options symbolgen mlogic mprint;
%let howmany = 1728; *This is the number of randomizations after applying stratification criteria;
%macro dup;
%do i = 1 %to &howmany;
data ds&i;
set r;
rand = &i;
```


J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.

```
output;
run;
%end;
%mend dup;
%dup quit;
%macro names(howmany,dataname);
%do i = 1 %to &howmany;
&dataname&i
%end;
%mend names;
data final;
set %names (&howmany,ds);
run;
```

5. Merge Cluster-Level Data with Randomization File

```
data sasfiles.randckd2;
   merge sasfiles.rand18a final;
   by rand;
   run;
```

6. Assign Clusters to Control or Intervention Group for Each Randomization

```
data sasfiles.randckd2;
   set sasfiles.randckd2;
   group = 1;
   if col1 = id or col2 = id or col3 = id or col4 =
id or col5 = id or col6 = id
   or col7 = id or col8 = id or col9 = id then
group = 2;*these were selected by proc IML in
each randomization and are assigned to group 1;
*the remaining will be assigned to group 2;
   run;
```

7. Create Group X Randomization Variable

```
data sasfiles.randckd2;
   set sasfiles.randckd2;
   randgrp = rand*100+group;
   run;
```

8. Output Group X Randomization Variable Means

```
proc summary data = sasfiles.randckd2 mean;
   by randgrp;
   var rand group var1 var2 var3 var4;
   output out = sasfiles.randckdsum;
   run;
```

9. Compute Sum of Squared Difference Between Groups on Standardized Variables for Each Randomization and Compute Balance Criterion Variable

Remember that var1 through var4 are now *z* scores of the original variables.

```
data sasfiles.randckdsum1;
   set sasfiles.randckdsum;
   by rand;
   retain dsvar1 dsvar2 dsvar3 dsvar4;
   if first.rand then dsvar1 = var1;
   if first.rand then dsvar2 = var2;
   if first.rand then dsvar3 = var3;
   if first.rand then dsvar4 = var4;
   *We want the absolute value of differences in-
stead of raw differences for later purposes (to com-
pare optimal versus remaining randomizations;
squared differences will be the same);
   if not first.rand then dsvar1 = abs(dsvar1-var1);
   if not first.rand then dsvar2 = abs(dsvar2-var2);
   if not first.rand then dsvar3 = abs(dsvar3-var3);
   if not first.rand then dsvar4 = abs(dsvar3-var4);
   if last.rand then output;
   run;
   /*get sums of squared differences between
groups*/
   data sasfiles.randckdsum2;
   set sasfiles.randckdsum1;
   sqdvar1 = dsvar1**2;
   sqdvar2 = dsvar2**2;
   sqdvar3 = dsvar3**2;
   sqdvar4 = dsvar4**2;
   totalsqd = sqdvar1 + sqdvar2 + sqdvar3 +
sqdvar4; *this is the balance criterion variable;
   keep rand totalsqd dsvar1 dsvar2 dsvar3 dsvar4;
   run;
```

10. Examine the Frequency Distribution of the Balance Criterion and Choose Cutpoint for Optimal Set

```
proc univariate data = sasfiles.randckdsum2;
   var totalsqd;
   run;
   *assign a random number to select final randomiza-
tion from optimal set (identified by variable bestrand;
   data sasfiles.randckdbest;
   set sasfiles.randckdsum2;
   bestrand = 0;
   x = rand('uniform');
   if totalsqd<1.445 then bestrand = 1;
   keep rand totalsqd bestrand x;
   run;
```


J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.

11. Choose Final Randomization for Study (Usually Based on the Lowest or Highest Value of x in the Optimal Set)

County-Level Data for Study 1

We obtained baseline information on the counties from the Colorado Department of Public Health and Environment, the CIIS, and 2010 Census data. The county-level variables we used are listed below:

1. Percentage of children between 0 and 4 who had ≥2 immunization records in CIIS
2. Number of 19- to 35-month-olds by county
3. Up-to-date rate (percentage) at baseline
4. Population by race (white, black, or other), ethnicity (Hispanic/Latino or not), and income
5. Ratio of pediatric to family medicine practices and number of community health centers

	Raw County-Level Variables for Study 1								
County	Location	In CIIS (%)	Children Ages 19–35 Months (n)	Up to Date on Immunizations (%)	African American Race (%)	Hispanic Ethnicity (%)	Average Income ($)	Pediatric Practice—to–Family Medicine Practice Ratio	Community Health Centers (n)
1	Rural	94	366	37	2	44	35,988	1.00	1
2	Rural	85	1,274	39	0	23	67,565	0.08	0
3	Rural	85	614	42	5	12	35,879	0.33	3
4	Rural	93	1,720	39	1	18	63,617	0.33	6
5	Rural	82	242	31	1	6	59,118	0.20	0
6	Rural	80	350	27	3	15	57,179	0.00	3
7	Rural	94	401	49	1	38	29,738	0.20	3
8	Rural	103	234	37	1	39	37,350	0.00	1
9	Urban	93	3,779	51	4	35	52,923	0.15	11
10	Urban	89	11,807	51	10	17	58,302	0.45	6
11	Urban	83	9,453	54	2	7	93,819	0.61	1
12	Urban	70	12,354	29	8	13	54,839	0.26	10
13	Urban	93	10,008	50	2	13	63,857	0.34	3
14	Urban	85	5,343	36	2	10	53,502	0.18	7
15	Urban	82	3,143	38	3	39	39,570	0.27	7
16	Urban	84	6,056	43	1	28	52,457	0.10	8

J Am Board Fam Med: first published as 10.3122/jabfm.2015.05.150001 on 9 September 2015. Downloaded from http://www.jabfm.org/ on 28 March 2024 by guest. Protected by copyright.