



Application of randomization techniques for balancing site covariates in the adult day service plus pragmatic cluster-randomized trial

David L. Roth^{a,*}, Jin Huang^a, Laura N. Gitlin^b, Joseph E. Gaugler^c

^a Center on Aging and Health, Division of Geriatric Medicine and Gerontology, 2024 East Monument Street, Johns Hopkins University, Baltimore, MD, 21205, USA

^b College of Nursing and Health Professions, 1601 Cherry Street, Mail Stop 10501, Drexel University, Philadelphia, PA, 19102, USA

^c Division of Health Policy and Management, School of Public Health, 420 Delaware St. SE, University of Minnesota, Minneapolis, MN, 55455, USA

ARTICLE INFO

Keywords:

Cluster-randomized trials
Pragmatic trials
Re-Randomization
Covariate-constrained randomization
Half-normal distribution
Family caregiving

ABSTRACT

Cluster-randomized trials (CRTs) are increasingly common in pragmatic trials of interventions for older adults, where staff of existing clinics or service agencies deliver interventions. The Adult Day Service (ADS) Plus intervention is delivered by trained staff at adult day service facilities to assist older adults with cognitive impairments and their family caregivers. Because sizable imbalances on important site characteristics might emerge from a simple randomization, we implemented a 3-stage constrained randomization approach to limit imbalance between intervention and usual care control conditions on 5 site characteristics: capacity; % of minority clients; % of clients with dementia; urban, rural or suburban location; and private or public ownership. In stage 1, the Balance Match Weighted (BMW) re-randomization procedure was used to assign 30 sites to ADS Plus or control arms based on the best randomization out of 20 total randomizations for minimizing site imbalance. In stage 2, propensity scores from the BMW logistic regression analysis for reserve sites were used to determine substitutions for randomized sites that opted out of the CRT prior to implementation. In stage 3, a minimization approach was used to add 20 more sites to the trial. A standardized metric based on the half-normal distribution of the absolute value of mean differences was used to assess site imbalance. After stage 3, the remaining imbalance for the 49 enrolled sites was reduced by 75% from what would have been expected from a simple randomization. Optimized randomization procedures with similar imbalance metrics should be used more routinely in pragmatic CRTs.

1. Introduction

Cluster-randomized trials (CRTs) allocate existing structural sites (e.g., hospitals, clinics, residential facilities) or social units to intervention or control conditions. CRT designs are increasingly common, particularly in pragmatic trials where site staff are trained to deliver interventions to groups of eligible patients or clients. In many CRTs, the number of sites or clusters to be randomized is relatively small, which can lead to important imbalances between intervention and control conditions on site-related characteristics [1–3]. Imbalance on either site or individual participant covariates can result in substantial bias when estimating treatment effects [4]. For this reason, many investigators implement optimized randomization procedures to reduce imbalances on key site characteristics in CRTs.

In this report, we describe the implementation of multiple optimized randomization procedures to minimize and control site characteristic imbalance in the Adult Day Service (ADS) Plus CRT. ADS Plus is an intervention delivered at adult day service facilities for older adults with dementia and other serious impairments. An important goal of adult day services is to provide respite to family members from caregiving responsibilities [5,6], and ADS Plus augments these services by providing caregivers with additional support, education, and care management skills [7,8]. The ADS Plus CRT is a pragmatic trial that uses a hybrid design [9] to test intervention effectiveness and implementation challenges. The goals and methods of the ADS Plus CRT have been described in detail previously [10]. In this paper, we carefully describe the implementation of the randomization procedures that were used.

The original study design called for 30 ADS sites from across the

Abbreviations: ADS, Adult Day Service; AVDM, Absolute Value of the Standardized Difference in the Mean; BMW, Balance Match Weighted; CRT, Cluster-Randomized Trial; SD, Standard Deviation; SQRT, Square Root.

* Corresponding author.

E-mail addresses: droth@jhu.edu, droth@jhu.edu (D.L. Roth), j.huang@jhu.edu (J. Huang), lng45@drexel.edu (L.N. Gitlin), gaug0015@umn.edu (J.E. Gaugler).

<https://doi.org/10.1016/j.conctc.2020.100628>

Received 24 February 2020; Received in revised form 14 July 2020; Accepted 26 July 2020

Available online 28 July 2020

2451-8654/© 2020 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

country to participate in the trial with an average of 10 participants enrolled at each site. A 1:1 site allocation plan was proposed, with 1/2 of the sites' staff receiving training to deliver the ADS Plus program and 1/2 serving as usual care control sites. An important goal was to control possible imbalance between randomized sites on 5 important site characteristics: ownership type (public vs. private); location (urban, suburban, or rural); site size or client capacity; % of clients who were White, non-Hispanic; and % of clients with a dementia diagnosis.

Optimized random allocation procedures for controlling site imbalance in CRTs can be categorized into four general types: matching, stratification, minimization, and covariate-constrained randomization [1]. Covariate-constrained randomization comprises a category of different approaches that share the following features: 1) Site data on key site characteristics are available to investigators before randomization occurs; 2) multiple random assignments of these sites to intervention conditions are conducted; 3) the degree of imbalance on the selected site covariates after each randomization is examined; and 4) an optimal or acceptable random assignment is selected from the multiple randomizations examined. Because multiple possible randomizations are evaluated by design, these approaches are sometimes referred to as *re-randomization* [11]. Specific covariate-constrained re-randomization approaches include methods developed by Moulton [12] and the Balance Match Weighted (BMW) approach designed by Xu and Kalbfleisch [13]. In the ADS Plus CRT, a combination of BMW and minimization procedures were implemented sequentially to control site imbalance on the site characteristics while still maintaining the experimental rigor of a randomized trial design.

2. Methods

2.1. Site recruitment and selection

Initially, 73 ADS sites expressed an interest in participating in the trial. The director of each site was contacted and basic descriptive data about the sites were systematically collected. This included information on the site size or client capacity, the race/ethnicity of site clients, the proportion of site clients with a dementia diagnosis, the location of the site (urban, suburban, or rural), and whether the site was privately or publicly owned. Of the 73 interested sites, 44 were deemed eligible to participate (e.g., of sufficient size, with adequate staff).

The original randomization protocol called for 30 sites to be included in this CRT, with 15 sites randomly assigned to the ADS Plus condition and 15 sites serving as usual care control sites. Among the 44 eligible sites, 11 were specifically selected to be included in the trial based on goals to include certain under-represented site types (e.g., suburban location, publicly-owned). Of the remaining 33 eligible sites, 19 were randomly selected by the study biostatistician to be included in the original random allocation using a random number generator provided by the SAS system. The remaining 14 sites were held in reserve.

It was specified before the trial began that acceptable balance was desired on 5 site characteristics (site capacity; % of minority clients; % of clients with dementia; urban, rural or suburban location; and private or public ownership). Because location had 3 levels (urban, suburban, or rural), this characteristic was represented by two dummy-coded variables, with urban as the reference. Consequently, a total of 6 balancing variables were used in the analyses. All analyses were conducted using version 9.4 of the SAS statistical software package [14].

2.2. Measuring imbalance across site characteristics

When considering different optimized randomization approaches for CRTs, suitable metrics of covariate imbalance and conventional benchmarks on what is considered to be acceptable balance would be useful [15]. Raab and Butcher [16] proposed a balance criterion (B) that consisted of a weighted sum of squared mean differences for site variables. They showed that B was positively skewed but offered little other

information on its statistical properties or reasonable thresholds for controlling imbalance. Consistent with previous recommendations [16], the weights used in the present study consisted of the inverse of the variance of the difference in means for each variable. Specifically, B was calculated as follows:

$$B = \sum_{i=1}^k w_i (\bar{x}_{0i} - \bar{x}_{1i})^2$$

where k denotes the number of site-level variables to be balanced and w_i is a pre-specified weight for the i th site-level variable.

An alternative to B is a recently developed standardized criterion for imbalance (H) that is based on the half-normal distribution of the absolute value of the standardized mean differences between intervention and control arms on the site covariates [17]. Briefly, when the absolute value of a normally distributed variable with a mean of 0 (such as a difference in means) is taken, the normal distribution is folded at that population mean ($\mu = 0$) and the absolute value is distributed as the positively-skewed, half-normal distribution [18,19]. In calculating H, the observed mean difference between sites (intervention – control) for each randomization is obtained and standardized ($M = 0, S = 1$) for each characteristic on which balance is desired. Next, the absolute values of those standardized mean differences (AVDM) are determined. For any one site characteristic, under the half-normal distribution the expected value of AVDM is 0.80 and its variance is 0.36. When the mean of the AVDMs is calculated across k different balancing covariates for each randomization, the central limit theorem dictates that this quantity (H) begins to approach the normal distribution with a mean 0.80 and a variance of $0.36/k$ (as k becomes large).

H, therefore, was calculated as follows.

$$H = \frac{1}{k} \sum_{i=1}^k \frac{1}{S_{DMi}} |\bar{x}_{0i} - \bar{x}_{1i}|$$

where S_{DMi} is the standard deviation of the difference in means for the i th site-level variable. H should approximate 0.80 for any simple randomization, and the extent to which a randomization results in an H less than 0.80 will reflect the degree of reduced imbalance or improved balance achieved by that randomization. With $k = 6$ in the ADS Plus CRT, the variance and standard deviation of H are expected to be 0.06 and 0.24, respectively.

2.3. Site enrollment in the ADS plus CRT

Sites were enrolled and assigned to the ADS Plus or control conditions across three distinct stages in the trial. These stages are described in detail below. A visual summary of the sites enrolled and maintained at each stage is displayed in Fig. 1.

2.4. BMW randomization (stage 1)

The BMW approach [13] was used to assign the original 30 sites to ADS Plus or control conditions. This approach involves conducting a pre-specified number of randomizations, calculating propensity scores of intervention assignment as a function of the covariates to be balanced, matching intervention and control sites on those propensity scores, and then selecting the one randomization that minimizes the propensity score differences of the matched sites. Unlike many other covariate-constrained re-randomization approaches that conduct either all or many possible randomizations and then randomly select one from a subset of randomizations with acceptably low imbalance [e.g., 12], the BMW approach typically specifies fewer possible randomizations *a priori* and then selects the optimal one from those examined. In the ADS Plus trial, we conducted 20 different simple randomizations of the 30 initial sites into 15 treatment and 15 control condition sites. Because there are over 77 million possible unique allocations of 15 sites to each

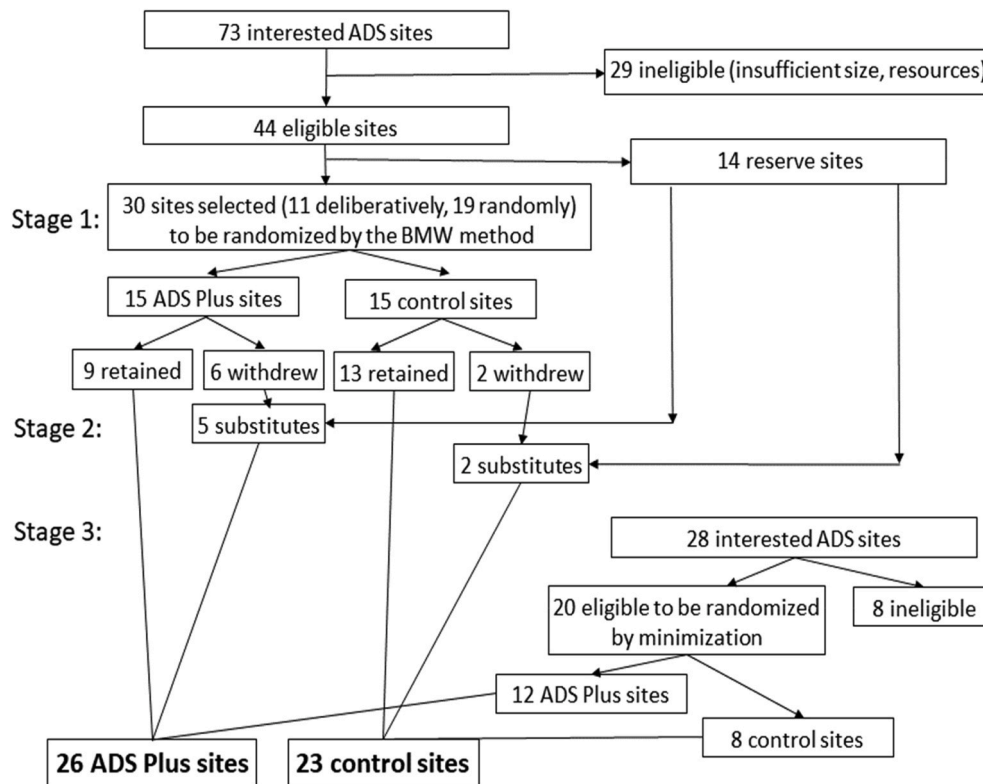


Fig. 1. Assignment and retention of sites across the three stages of the ADS plus cluster-randomized trial.

intervention arm, the 20 randomizations conducted constitutes only 0.0000003 of all possible allocations. For each of these 20 randomizations, the following steps were executed:

1. Using a binary logistic regression model, the actual intervention assignment (ADS Plus vs. usual care) was regressed on the 6 balancing covariates.
2. For each randomized site, the propensity score for ADS Plus assignment was calculated based on that site's scores on the balancing covariates and the weights from the logistic regression model.
3. Intervention and control sites were then matched on these propensity scores as closely as possible, allowing up to two sites in one treatment arm to be matched to any single site in the other treatment arm.
4. The overall difference in propensity scores was calculated across the matched sites.

The parameters of this method (e.g., number of randomizations, up to 2:1 matching of sites by propensity scores) were specified *a priori*. The randomization that resulted in the smallest overall difference in propensity scores was selected as the optimal randomization from among the 20 randomizations examined. Although decisions were made based on propensity score differences, both B [16] and H [17] were also examined as imbalance criteria.

2.5. Substitution (stage 2)

The 44 initially eligible sites expressed a willingness to participate in the trial several months before funding was secured and site randomization began. Due to changes in site leadership and other organizational factors, some sites ($n = 7$) withdrew after the BMW randomization but before any staff training procedures were implemented or individual participants were enrolled. This included 5 sites assigned to ADS Plus and 2 sites assigned to the usual care control condition.

For the 14 sites that were held in reserve after the BMW

randomization, the propensity scores for ADS plus assignment were calculated based on the logistic regression equation from the optimal BMW randomization. For each sequential site that withdrew, the reserve site with the closest propensity score to that of the withdrawing site was selected and substituted for that withdrawing site. The imbalance criteria B and H were re-calculated after each such substitution to track the control of site imbalance.

2.6. Minimization (stage 3)

After over a year of participant recruitment within randomized sites, it became clear that several sites were having difficulty enrolling up to 10 participants. One ADS Plus site had failed to enroll any participants and became the 8th site to withdraw from the trial. This led senior investigators to recruit additional sites for possible inclusion in the trial in order to make timely progress toward the original goal of enrolling 300 individual participants in the trial. A total of 28 new sites communicated a willingness to participate in the ADS Plus CRT, and 20 of these sites were deemed to be eligible.

Site information was obtained from these 20 new sites, and those sites were then randomly ordered by the study statistician using a SAS random number generator. Next, for each sequential site on this randomly ordered list, we calculated the H metric that would result if the site was assigned to the ADS Plus condition or to the control condition. The site was then assigned to that condition which resulted in the lower H metric. Because sites assigned to ADS Plus appeared to be more likely to drop out, the investigators decided to continue this minimization selection process until 12 sites had been assigned to the ADS Plus condition and 8 sites assigned to the control condition.

3. Results

3.1. BMW randomization (stage 1)

Descriptive data on the site balancing covariates for the 30 sites randomized in stage 1 are presented in [Table 1](#). Descriptive data from the optimal BMW randomization are also presented in [Table 1](#). Based on the half-normal distribution, the mean (across the 20 BMW randomizations) of the AVDMs for each balancing variable should be approximately equal to its expected value of 0.80 and the SD of this mean (across the 20 BMW randomizations) should be 0.13 (SQRT(0.36)/SQRT(20)). As the 3rd column of [Table 1](#) indicates, 4 of the 6 AVDMs are within 1 SD of 0.80, and the overall mean of these means (0.82) is also within the sampling distribution of the expected mean of 0.80.

Descriptive information for the optimal BMW randomization is also summarized in [Table 1](#), and the imbalance metrics observed for each of the 20 randomizations that were conducted as part of the BMW approach are presented in [Table 2](#). The 13th randomization was chosen to be the optimal randomization based on the lowest difference in matched site propensity scores. The results in [Table 1](#) indicate that the AVDM for each balancing variable from this optimal randomization was below the expected value of that metric of 0.80 for a typical simple randomization. Because the half-normal distribution is not symmetric, the median is not equal to the mean and has an expected value of 0.67 [17]. As indicated in [Table 1](#), 5 of the 6 balancing variables have AVDMs from the optimal BMW randomization that were also below the expected value for the median. For the % of clients with a dementia diagnosis, the AVDM was 0.79, which is above the expected value for the median and corresponds to the 57th percentile.

The H metric is the overall mean of the AVDMs across the 6 balancing covariates for each randomization [17]. For the optimal randomization, $H = 0.34$, a value that is 1.92 SDs below the expected mean given by the half-normal distribution and at the 3rd percentile. It also represents a 58% improvement in balance achieved between the expected balance from a single, simple randomization (0.80) and perfect covariate balance (0.00). The percentile for the H metric reported in [Table 2](#) is based on the theoretical distribution of H when $k = 6$.¹ Although the optimal randomization was chosen based on the lowest difference in matched site propensity scores from the BMW method, [Table 2](#) indicates that the same choice of would have been made based on picking the randomization with the lowest B or H. Correlational analyses indicated that the propensity score difference, B, and H were all highly inter-correlated (r 's range from 0.69 (propensity diff with B) to 0.96 (B with H)).

3.2. Site withdrawals and substitutions (stage 2)

[Table 3](#) reports the results of the site-by-site substitutions as part of stage 2. Both imbalance metrics (B and H) showed steady increases as these substitutions progressed. The final H of 0.72 after 7 BMW propensity-score based substitutions was still better than the benchmark of 0.80 from a single, simple randomization, but much of the balancing advantages of the BMW approach were found to dissipate after these site substitutions.

3.3. Sites added by minimization (stage 3)

[Table 4](#) reports the results on the site imbalance metrics B and H after the 20 new sites were randomly ordered and then assigned based on the

minimization method. As expected and by design, H decreased sequentially, as did B. The final H of 0.20 is 2.50 SDs below the expected mean of 0.80 and at the 1st percentile. With an expected imbalance of 0.80 from a simple randomization, $H = 0.20$ represents a 75% reduction in actual imbalance on the site characteristics.

At the conclusion of the 3-stage site allocation process, 32 sites had been assigned to ADS Plus and 25 to control. All assignments had a random component and no site was preferentially positioned for assignment to either intervention condition. Six sites assigned to ADS Plus and two sites assigned to control dropped out before enrolling any participants, resulting in 26 retained ADS Plus sites and 23 control sites. However, before many of the stage 3 sites could be trained and participants enrolled, the COVID-19 pandemic struck. Most ADS sites suspended their on-site activities and participant enrollment into the ADS Plus CRT was suspended as well. At this point in March of 2020, we had 32 active sites that had enrolled participants at that time. [Table 5](#) presents descriptive data on the site characteristics of the 49 retained sites and on subset of 32 active sites. All AVDM's for individual balancing covariates are substantially smaller than 0.80 for the 49 retained sites. For the 32 active sites, the AVDM for race approaches the $H = 0.80$ benchmark and the AVDM for % of clients with a dementia diagnosis actually exceeds this benchmark, indicating some imbalance on that site characteristic that is, fortunately, still not statistically significant. Overall, the H of 0.53 (14th percentile) for the active sites indicated adequate control of site imbalance across the site characteristics collectively.

4. Discussion

This paper illustrates the random allocation methods that were used in the pragmatic, hybrid CRT for the ADS Plus intervention. We originally chose the BMW method [13] to minimize site imbalance on 5 site characteristics by selecting the best of 20 simple randomizations of 30 original sites to 15 ADS Plus and 15 control sites. Although the propensity-score differences of matched sites was the criterion for identifying the optimal randomization, the same choice would have been made had either B [16] or H [17] been used. Thus, some of the steps of the BMW approach (calculating propensity scores from logistic regression, matching individual sites on those propensity scores) could have been avoided had we just proceeded with using either of these other imbalance metrics. Although the BMW propensity scores were also used in stage 2 when we selected substitute sites for those who withdrew, the H metric indicated that some degree of imbalance returned as this substitution process continued. After seven such site substitutions, H approached the 0.80 expected value from just one simple randomization.

After 20 additional sites were added using a minimization method based on H, excellent balance between intervention conditions on site characteristics was re-established. Site imbalance across all site characteristics was generally controlled for the 32 active sites, although some non-significant imbalance is present for the proportion of site clients who have a diagnosis of dementia. Our goal throughout this pragmatic trial was to maintain the experimental rigor of randomization while also minimizing random imbalance between ADS Plus and control sites on several important site characteristics. We had to adapt to multiple site withdrawals after our original randomization but before intervention implementation and incorporate procedures that allowed us to add additional sites after the original randomization. Sites dropped out for multiple reasons including changes in administration and resources. Such changes can be expected in pragmatic trials involving community-based programs, especially when several months if not years pass between site recruitment for a grant submission and subsequent funding of the project. The end result was the identification of 49 retained sites, including 32 sites that were fully trained and actively enrolling participants before the suspension of new enrollments due to the COVID-19 pandemic.

¹ Because $k = 6$ may not be considered sufficiently large, we ran a simulation with 10,000 trials for the initial 30-site randomization and determined an empirical standard deviation of 0.28, an empirical median was 0.79, and empirical percentiles for H. Because these values were very close to the theoretical values given by the central limit theorem, the empirical or simulated values are not further reported in the paper.

Table 1
Characteristics of the Original 15 ADS Plus and 15 Controls Sites Randomly Assigned with the BMW approach.

Characteristic	Overall mean	Overall SD	Mean AVDM across 20 randomizations	Optimal BMW Randomization				
				Mean (ADS Plus)	SD (ADS Plus)	Mean (Control)	SD (Control)	AVDM
Site Size	65.00	31.47	0.84	64.80	35.55	65.20	28.06	0.03
Public Ownership (Y = 1, N = 0)	0.17	0.38	0.82	0.20	0.41	0.13	0.35	0.48
Location1 (Suburban = 1, Other = 0)	0.17	0.38	0.87	0.13	0.35	0.20	0.41	0.48
Location2 (Rural = 1, Other = 0)	0.33	0.48	0.57	0.33	0.49	0.33	0.49	0.00
Race (% White)	57.04	36.44	0.85	55.11	37.89	58.97	36.16	0.28
Dementia Diagnosis %	79.63	19.87	0.96	82.53	17.42	76.73	22.29	0.79
Mean	-	-	0.82	-	-	-	-	0.34

Notes: ADS = Adult Day Service; AVDM = Absolute Value of the Standardized Difference in Means; BMW = Balance Match Weighted.

Table 2
Site balancing statistics by individual randomizations of the BMW approach.

Randomization	Difference in Matched Site Propensity Scores	B	H	Percentile for H (based on normal distribution)	Min AVDM	Max AVDM
1	2.46	4.65	0.77	45	0.20	1.47
2	3.13	7.83	0.88	63	0.00	2.01
3	0.66	2.39	0.50	11	0.00	1.08
4	0.52	1.50	0.39	5	0.00	0.93
5	2.46	4.65	0.77	45	0.20	1.47
6	2.28	7.64	0.98	77	0.23	1.56
7	3.32	12.49	1.15	92	0.48	2.92
8	3.72	4.44	0.77	45	0.34	1.47
9	3.79	17.28	1.48	100	0.48	3.11
10	3.07	11.68	1.25	97	0.00	2.09
11	3.34	10.76	1.06	86	0.23	2.69
12	3.07	11.68	1.25	97	0.00	2.09
13*	0.32	1.16	0.34	3	0.00	0.79
14	2.98	3.47	0.73	39	0.48	1.08
15	1.75	1.58	0.45	8	0.01	0.76
16	1.74	5.90	0.75	42	0.00	1.47
17	1.21	1.35	0.43	7	0.05	0.76
18	1.20	6.03	0.95	73	0.48	1.47
19	2.98	3.47	0.73	39	0.48	1.08
20	1.50	6.21	0.78	47	0.00	1.85

Notes: * denotes the selected randomization; AVDM = Absolute Value of the Standardized Difference in Means; BMW = Balance Match Weighted.

Table 3
Site balancing statistics after site substitutions (stage 2).

Substitution	Propensity Score of Site Dropped	Propensity Score of Site Substituted	B	H
0	-	-	1.16	0.34
1	0.44	0.46	2.04	0.35
2	0.41	0.48	2.01	0.44
3	0.33	0.48	2.71	0.54
4	0.42	0.50	2.56	0.47
5	0.54	0.54	2.40	0.50
6	0.46	0.51	2.64	0.49
7	0.40	0.52	3.96	0.72

Covariate-constrained randomization techniques are increasingly popular in CRTs, but uncertainties remain over the specifics of implementation, how overall imbalance should be assessed, and what constitutes acceptable balance thresholds [15]. Stratification is sometimes used in CRTs [20], but this method can become unwieldy when balance is desired on more than just a few site characteristics [1,3]. A classic method of covariate-constrained randomization was described by Moulton [12] and consisted of specifying, in advance, the minimal degree of imbalance desired on each specific covariate, conducting all possible randomizations, identifying the subset of possible randomizations that achieve acceptable control of imbalance, and randomly selecting one random allocation from that subset of acceptable randomizations. However, even a relatively modest number of sites can lead to an unwieldy number of total possible randomizations using this

Table 4
Site balancing statistics after assigning additional sites with minimization method (stage 3).

Site Added	Group Assignment	B	H
0 (after 7 site substitutions)		3.96	0.72
0 (after 1 substitute site dropped out)		5.98	0.86
1	ADS Plus	5.95	0.83
2	ADS Plus	3.61	0.74
3	ADS Plus	3.44	0.71
4	ADS Plus	3.05	0.66
5	ADS Plus	2.68	0.57
6	Control	1.64	0.48
7	ADS Plus	1.01	0.38
8	ADS Plus	1.00	0.37
9	Control	0.64	0.28
10	ADS Plus	0.67	0.29
11	Control	0.58	0.26
12	Control	0.64	0.27
13	Control	0.69	0.27
14	ADS Plus	0.58	0.27
15	Control	0.53	0.21
16	ADS Plus	0.60	0.25
17	Control	0.46	0.23
18	ADS Plus	0.27	0.18
19	ADS Plus	0.27	0.16
20	Control	0.13	0.20

method, and questions remain on how “acceptable balance” should be defined. Moulton [12] illustrated that sometimes study-specific covariate criteria need to be tightened or relaxed at interim stages and openly

Table 5

Characteristics of the 49 retained sites and 32 active sites.

Characteristic	Retained (26 ADS Plus sites and 23 Control sites)						Active (14 ADS Plus sites and 18 Control sites)					
	ADS Plus		Control		AVDM	P	ADS Plus		Control		AVDM	P
	Mean	SD	Mean	SD			Mean	SD	Mean	SD		
Site Size	71.65	39.19	67.04	30.86	0.46	0.65	72.86	47.07	63.78	32.52	0.62	0.52
Public Ownership (Y = 1, N = 0)	0.08	0.27	0.09	0.29	0.12	0.90	0.14	0.36	0.11	0.32	0.26	0.80
Location1 (Suburban = 1, Other = 0)	0.50	0.51	0.52	0.51	0.15	0.88	0.43	0.51	0.39	0.50	0.22	0.83
Location2 (Rural = 1, Other = 0)	0.15	0.37	0.13	0.34	0.23	0.82	0.14	0.36	0.17	0.38	0.18	0.86
Race (% White)	68.58	30.75	67.35	32.70	0.14	0.89	70.73	34.51	61.61	34.74	0.74	0.47
Dementia Diagnosis % Mean (H)	72.52	29.04	73.30	22.94	0.11	0.92	84.46	20.23	75.72	21.30	1.18	0.25
	–	–	–	–	0.20	–	–	–	–	–	0.53	–

Notes: ADS = Adult Day Service; AVDM = Absolute Value of the Standardized Difference in Mean.

discussed how such direct manipulation can contribute to appearances that investigators have “rigged the outcome (p. 301)” or “manipulated the design to his or her advantage. (p. 304).”

Adherence to pre-specified standards and more general acceptance of reasonable benchmarks of site balance would constitute important advances in this important area of methodological development. In terms of metrics and targets, B and H both represent metrics of overall imbalance across multiple site characteristics. As shown in this illustration, B and H are highly correlated, but H provides additional benefits in terms of a standardized scale that can facilitate comparative descriptions across studies and in reference to what would be expected from a simple randomization. In addition to overall imbalance control targets across the site covariates collectively, the degree of imbalance observed on individual site characteristics might also be examined and considered. Ciolino and colleagues [15] conducted simulation studies and endorsed a threshold of having all p-values from nonparametric analyses of individual site characteristics being 0.30 or higher as indicative of acceptable balance. However, p-values are largely dependent on the number of sites being randomized, making this threshold easier to achieve with some notable imbalances still remaining when the number of sites to be randomized is relatively small.

Optimized randomization procedures such as covariate-constrained randomization may have implications for the analysis of treatment effects. In most cases, the increased balance on the covariates from re-randomization approaches should result in more precise estimates of treatment effects and could make traditional parametric analyses overly conservative [11]. Randomization tests can be conducted as an alternative or in sensitivity analyses. The implications for outcome analyses in pragmatic CRTs that use optimized randomization procedures should further be investigated in future methodological studies.

We recommend that future applications of optimized randomization procedures for CRTs set *a priori* targets for controlling imbalance and use standardized metrics such as H. Such targets might also be supplemented with realistic criteria for individual balancing factors. For example, one could specify that the AVDMs for individual site characteristics be below a pre-specified cut point, such as the 25th percentile, that p-values for intervention vs. control comparisons of individual balancing factors all exceed 0.30 [15], and that only randomizations with H below the 10th percentile be considered as acceptable. For the 6 balancing variables that represented the 5 site characteristics in the ADS Plus trial, the 10th percentile corresponds to H = 0.48, and 4 of the initial 20 randomizations would have been considered acceptable. Investigators may choose to keep randomizing until a particular randomization meets *a priori* thresholds, or conduct a great number of randomizations (e.g., 100 or more) and then randomly select one from the subset of those randomizations that achieve acceptable balance.

As Moulton [12] has pointed out, care should be taken to ensure that treatment allocation decisions in CRTs are not impacted by individual site preferences or by other interim decisions that might introduce an appearance of bias. In the ADS Plus CRT, all site allocation decisions were made by statisticians who had no contact with project sites or their

personnel. In general, as investigators hope to optimize designs in pragmatic CRTs, it will be essential to implement procedures that maintain experimental rigor and minimize opportunities for biased estimates of treatment effects.

Funding

This work was supported by a grant R01 AG049692 from the National Institute on Aging (NIA). This work was also supported in part by the NIA of the National Institutes of Health under Award Number U54AG063546, which funds the NIA Imbedded Pragmatic Alzheimer's Disease and AD-Related Dementias Clinical Trials Collaboratory (NIA IMPACT Collaboratory). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Drs. Gitlin and Gaugler co-lead the Dissemination and Implementation (D&I) Core of the NIA Impact Collaboratory.

References

- [1] N.M. Ivers, L.J. Halperin, J. Barnsley, J.M. Grimshaw, B.R. Shah, K. Tu, et al., Allocation techniques for balance at baseline in cluster randomized trials: a methodological review, *Trials* 13 (2012) 120.
- [2] F. Li, Y. Lokhnygina, D.M. Murray, P.J. Heagerty, E.R. DeLong, An evaluation of constrained randomization for the design and analysis of group-randomized trials, *Stat. Med.* 35 (2016) 1565–1579.
- [3] M.L. Shaffer, E.M.C. D'Agata, D. Habtemariam, S.L. Mitchell, Covariate-constrained randomization for cluster randomized trials in the long-term care setting: application to the TRAIN-AD trial, *Contemp. Clin. Trials Commun.* 18 (2020) 100558.
- [4] S. Yang, M.A. Starks, A.F. Hernandez, E.L. Turner, R.M. Califf, C.M. O'Connor, R. J. Mentz, K.R. Choudhury, Impact of baseline covariate imbalance on bias in treatment effect estimation in cluster randomized trials: race as an example, *Contemp. Clin. Trials* 88 (2020) 105775.
- [5] H. Dabelko-Schoeny, S. King, In their own words: participants' perceptions of the impact of adult day services, *J. Gerontol. Soc. Work* 53 (2) (2010) 176–192.
- [6] N.L. Fields, K.A. Anderson, H. Dabelko-Schoeny, The effectiveness of adult day services for older adults: a review of the literature from 2000 to 2011, *J. Appl. Gerontol.* 33 (2) (2014) 130–163.
- [7] L.N. Gitlin, K. Reeve, M.P. Dennis, E. Mathieu, W.W. Hauck, Enhancing quality of life of families who use adult day services: short- and long-term effects of the adult day services plus program, *Gerontol.* 46 (5) (2006) 630–639.
- [8] K.E. Reeve, E. Mathieu, M.P. Dennis, L.N. Gitlin, Adult Day Services Plus: augmenting adult day centers with systematic care management for family caregivers, *Alzheimer Care Today* 4 (4) (2004) 332–339.
- [9] G.M. Curran, M. Bauer, B. Mittman, J.M. Pyne, C. Stetler, Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact, *Med. Care* 50 (3) (2012) 217–226.
- [10] L.N. Gitlin, K. Marx, D. Scerpella, H. Dabelko-Schoeny, K.A. Anderson, J. Huang, et al., Embedding caregiver support in community-based services for older adults: a multi-site randomized trial to test the Adult Day Service Plus Program (ADS Plus), *Contemp. Clin. Trials* 83 (2019) 97–108.
- [11] K.L. Morgan, D.B. Rubin, Rerandomization to improve covariate balance in experiments, *Ann. Stat.* 40 (2) (2012) 1263–1282.
- [12] L.H. Moulton, Covariate-based constrained randomization of group-randomized trials, *Clin. Trials* 1 (3) (2004) 297–305.
- [13] Z. Xu, J.D. Kalbfleisch, Propensity score matching in randomized clinical trials, *Biometrics* 66 (2010) 813–823.
- [14] SAS, Version 9.4. Cary, NC.

- [15] J.D. Ciolino, A. Diebold, J.K. Jensen, G.W. Rouleau, K.K. Koloms, D. Tandon, Choosing an imbalance metric for covariate-constrained randomization in multiple-arm cluster-randomized trials, *Trials* 20 (1) (2019) 293.
- [16] G.M. Raab, I. Butcher, Balance in cluster randomized trials, *Stat. Med.* 20 (3) (2001) 351–365.
- [17] J. Huang, D.L. Roth, Using the Half Normal Distribution to Quantify Covariate Balance in Cluster-Randomized Pragmatic Trials, 2020. Manuscript under review.
- [18] F.C. Leone, L.S. Nelson, R.B. Nottingham, The folded normal distribution, *Technometrics* 3 (4) (1961) 543–550.
- [19] M. Tsagris, C. Beneki, H. Hassani, On the folded normal distribution, *Mathematics* 2 (2014) 12–28.
- [20] K.M. Margolis, A.L. Crain, A.R. Bergdall, M. Beran, J.P. Anderson, L.I. Solberg, P. J. O'Connor, J.M. Sperl-Hillen, P.A. Pawloski, J.Y. Ziegenfuss, D. Rehrauer, C. Norton, P. Haugen, B.B. Green, Z. McKinney, A. Kodet, D. Appana, R. Sharma, N. K. Trower, R. Williams, B.F. Crabtree, Design of a pragmatic cluster-randomized trial comparing telehealth care and best practice clinic-based care for uncontrolled high blood pressure, *Contemp. Clin. Trials* 92 (2020) 105939.