Journal of Clinical Epidemiology 65 (2012) 132−137

ORIGINAL ARTICLES

The ‘‘best balance’’ allocation led to optimal balance in cluster-controlled trials

Esther de Hoop[a,*], Steven Teerenstra[a], Betsie G.I. van Gaal[b], Mirjam Moerbeek[c], George F. Borm[a]

[a]*Department of Epidemiology, Biostatistics and HTA, 133, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands*
[b]*Scientific Institute for Quality of Healthcare, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands*
[c]*Department of Methodology and Statistics, Utrecht University, 3508 TC Utrecht, The Netherlands*

Accepted 23 May 2011; Published online 12 August 2011

Abstract

Objective: Balance of prognostic factors between treatment groups is desirable because it improves the accuracy, precision, and credibility of the results. In cluster-controlled trials, imbalance can easily occur by chance when the number of cluster is small. If all clusters are known at the start of the study, the ‘‘best balance’’ allocation method (BB) can be used to obtain optimal balance. This method will be compared with other allocation methods.

Study Design and Setting: We carried out a simulation study to compare the balance obtained with BB, minimization, unrestricted randomization, and matching for four to 20 clusters and one to five categorical prognostic factors at cluster level.

Results: BB resulted in a better balance than randomization in 13−100% of the situations, in 0−61% for minimization, and in 0−88% for matching. The superior performance of BB increased as the number of clusters and/or the number of factors increased.

Conclusion: BB results in a better balance of prognostic factors than randomization, minimization, stratification, and matching in most situations. Furthermore, BB cannot result in a worse balance of prognostic factors than the other methods.

Keywords: Cluster randomized trial; Imbalance; Prognostic factors; Minimization; Randomization; Stratification; Matching

In any trial, it is desirable to have a similar distribution of prognostic factors in all treatment groups [1−4]. A good balance of prognostic factors improves the accuracy and precision of the results and enhances the credibility and acceptance of the results [4−7]. Randomization is an often-applied method to obtain a good balance of prognostic factors. However, random allocation may produce substantial imbalances between treatment groups, especially in small trials or when there are many prognostic factors.

In cluster-controlled trials (CCTs), complete social units, or clusters of individuals (such as families or medical practices), are randomized over the treatment arms of a study [8]. Such trials are being used more and more frequently in health services research [9], mostly because of feasibility considerations or to prevent contamination [8].

Especially when the number of clusters is small, simply randomizing clusters over the treatment arms can easily result in unequally distributed cluster characteristics because

* Corresponding author. Tel.: +31-24-3667262.
E-mail address: E.deHoop@ebh.umcn.nl (E. de Hoop).

0895-4356 © 2012 Elsevier Inc. Open access under the Elsevier OA license.
doi: 10.1016/j.jclinepi.2011.05.006

iring) or stratification is often used to promote balance in CCTs. However, these methods have serious limitations when the number of clusters is small. Matching results in a loss of efficiency [10], and stratification can be used to balance only a limited number of prognostic factors [6].

An alternative allocation approach is minimization [11]. This method allocates subjects with any number of characteristics to treatment groups to make the groups most nearly balanced. Minimization has been shown to improve balance better than other allocation methods in individually randomized trials. In addition, it allows for balancing on more prognostic factors than, for example, stratification [3,12].

Minimization can also be used in CCTs. The assignment of a cluster by this method to a treatment group is originally conceived as sequential, so based on the order in which clusters enter. Each subsequent cluster will be assigned to the treatment arm that produces the least imbalance [1,11]. However, in CCTs, the clusters are often known at the start of the study, so the allocation of clusters to treatment groups can be done all at the same time. In this situation, we propose the best balance method (BB): a method that divides clusters over two groups in all possible ways and selects

What is new?

- We compare five important methods to obtain balance of prognostic factors in cluster-controlled trials with small numbers of clusters: the best balance allocation method, minimization, unrestricted randomization, matching, and stratification.
- The best balance allocation method results in a better balance of prognostic factors than randomization, minimization, matching, and stratification when all clusters are known in advance.
- In cluster-controlled trials with limited numbers of clusters, the best balance allocation method is a preferable method of treatment allocation.

an allocation scheme in which the groups are optimally balanced [7].

The objective of this article is to compare BB with unrestricted randomization, minimization, matching, and stratification in terms of achievable balance in CCTs with small numbers of clusters. Comparisons between the allocation methods will be illustrated with an example.

2. Methods

2.1. Measure of imbalance

An imbalance measure must be computed for each category of each factor. The category imbalance measure is the difference between the numbers of clusters in each treatment group that are in that category. This imbalance measure requires that the factors are categorical, so continuous factors should be categorized first. (Although minimization based on ranks is also possible for continuous factors [13,14].) The overall imbalance is then defined as the sum of the squared category imbalances. For example, suppose that the variable ''ward'' has two categories—surgical wards and internal medicine wards. If there are three surgical wards and two internal medicine wards in treatment group A and there are two surgical wards and three internal medicine wards in treatment group B, then the difference in surgical wards is $3-2=1$ and the difference in internal medicine wards is $2-3=-1$. If ''ward'' is the only factor, then the overall quadratic imbalance is $(1)^2+(-1)^2=2$. This type of imbalance measure is also used in the variance method of minimization [1,3].

2.2. Best balance method

The BB method calculates the imbalance for all possible allocation schemes. Then, the allocation schemes that show the least imbalance are selected. Finally, one of these schemes is randomly chosen for implementation [7] (the R-code of this procedure can be found at http://ebh-research.ruhosting.nl/Best-Balance/).

2.3. Simulation study

We performed a simulation study to compare BB with unrestricted randomization, minimization, and matching. Trial data sets were generated in which the number of clusters and the number of factors varied. The numbers of clusters were 4, 6, 8, 10, 16, and 20, and the number of factors ranged from one to five, with two categories per factor. For every number of factors, all possible configurations were obtained. For example, if we have two factors with categories coded A and B, then there are four configurations possible: AA, AB, BA, and BB. Next, configurations were randomly drawn (with replacement) from this set and assigned to the clusters, which resulted in a data set. Ten thousand−trial data sets were generated for every combination of number of clusters and number of factors.

BB, minimization, unrestricted randomization, and matching were then used to divide the clusters over two treatment arms in every trial data set. Minimization allocated the clusters one by one in the order of the data set. For matching, 10,000 random matchings were performed for every data set. Then, one of the best matching schemes was chosen at random. The quadratic imbalance was calculated for all methods and compared with the imbalance obtained with BB.

3. Results

The results of the simulation study are shown in Table 1. This table shows the percentage of data sets in which BB resulted in less imbalance than randomization, minimization, and matching, respectively. Because the BB method involves choosing from among the allocations with the lowest quadratic imbalance, it can never result in a higher imbalance than the other methods. So, in the remaining percentage of the data sets, the other methods showed the same imbalance as BB.

In the case of four clusters and one prognostic factor, BB resulted in less imbalance than randomization in 13% of the trial data sets. As the number of factors increased, the percentage of data sets in which BB resulted in less imbalance also increased, up to 42% for four clusters and five factors. The same trend was found for larger numbers of clusters. Additionally, it can be seen that as the number of clusters increased, BB resulted in less imbalance than randomization in an increasing percentage of data sets, even up to 100%.

Minimization and BB performed equally well if there is only one prognostic factor. However, as the number of factors increased, BB resulted in a better balance than minimization in 3−61% of the data sets. This result also depends

Table 1. Results of simulation study: percentage of situation in which BB shows less imbalance than randomization, minimization, and matching

N clusters	*N* factors	BB shows less imbalance than randomization[a]	BB shows less imbalance than minimization[a]	BB shows less imbalance than matching[a]
4	1	13	0	0
	2	25	0	4
	3	32	9	8
	4	38	15	12
	5	42	20	16
6	1	23	0	0
	2	40	3	6
	3	54	16	18
	4	61	29	29
	5	67	37	37
8	1	74	0	0
	2	92	4	7
	3	98	19	20
	4	99	35	34
	5	100	48	46
10	1	100	0	0
	2	100	4	5
	3	100	19	23
	4	100	38	43
	5	100	52	57
16	1	100	0	0
	2	100	6	12
	3	100	22	43
	4	100	38	65
	5	100	57	79
20	1	100	0	0
	2	100	8	28
	3	100	25	58
	4	100	40	76
	5	100	61	88

Abbreviation: BB, best balance.
[a] BB will never result in a higher imbalance.

on the number of clusters. The effect became larger as the number of clusters increased.

The comparison between matching and BB produced essentially the same results as the comparison between minimization and BB. As the number of prognostic factors increased, the percentage of data sets in which BB shows less imbalance than matching also increased. This effect became larger as the number of clusters increased. The results for minimization and matching were almost the same with up to eight clusters. However, with more than eight clusters, matching performs worse in more data sets than minimization, both in comparison to BB.

Table 2 shows the median and maximum quadratic imbalances of the trial data sets. The median imbalances of the data sets are almost the same for BB, minimization, and matching. Only with larger number of clusters and factors, the medians of minimization and matching are higher than those of BB. In general, the medians for randomization are substantially larger than those of BB.

The maximum quadratic imbalance in the trial data sets was compared as well. BB showed a smaller maximum quadratic imbalance than the other allocation methods. Especially with increasing numbers of factors and clusters, the maximum imbalance resulting from randomization became very high. The same effect was found for matching, although to a lesser extent than for randomization. The maximum imbalance of minimization was just somewhat higher than the maximum imbalance found for BB.

4. Example: Safe or Sorry? study

Patients in hospitals and nursing homes are at risk of developing often-preventable adverse events, which threaten patient safety. van Gaal et al. [15] developed an integral patient safety program that addresses several adverse events simultaneously. These events are pressure ulcers, falls, and urinary tract infections. The program was tested in a CCT in 10 hospital wards and 10 nursing home wards. The primary outcome measure was the incidence of adverse events on every ward. Patient characteristics, length of stay, and nurse characteristics differed between hospitals and nursing homes; so it was decided before the start of the study that the results would be analyzed separately.

Table 2. Median and maximum quadratic imbalances of the generated data sets obtained by BB, randomization, minimization, and matching

		BB		Randomization		Minimization		Matching	
***N* clusters**	***N* factors**	**Median**	**Maximum**	**Median**	**Maximum**	**Median**	**Maximum**	**Median**	**Maximum**
4	1	1	2	2	8	1	2	1	2
	2	2	4	2	16	2	4	2	8
	3	4	8	4	24	4	10	4	10
	4	4	10	7	26	6	12	4	16
	5	6	12	8	28	6	18	6	18
6	1	2	2	2	18	2	2	2	2
	2	2	4	4	36	2	8	2	8
	3	2	8	8	34	4	12	4	20
	4	4	10	12	46	6	20	6	22
	5	6	12	14	54	8	24	8	30
8	1	0	2	4	34	0	2	0	2
	2	2	4	8	54	2	10	2	8
	3	4	8	16	64	4	12	4	20
	4	4	10	20	70	6	26	6	36
	5	6	12	30	82	10	30	8	38
10	1	0	2	10	58	0	2	0	2
	2	2	4	24	84	2	8	2	8
	3	4	6	36	100	4	18	4	20
	4	4	8	49	98	6	30	6	36
	5	5	10	62	132	10	28	12	44
16	1	0	2	10	80	0	2	0	2
	2	2	4	26	122	2	10	2	18
	3	2	6	44	162	4	20	6	32
	4	4	8	59	254	6	24	12	58
	5	4	10	78	190	10	36	16	60
20	1	2	2	10	106	2	2	2	2
	2	2	4	32	138	2	8	4	20
	3	4	6	50	162	4	20	10	52
	4	4	8	68	252	6	24	14	62
	5	4	10	86	230	10	34	22	106

Abbreviation: BB, best balance.

For the sake of simplicity, we will focus on the hospitals only. The program was tested in four internal medicine wards and six surgical wards. It was expected that the incidence of adverse events would differ per type of ward. Therefore, balancing on this factor was recommended. Other prognostic factors were the percentage of patients who were at risk of falling per ward ($<20\%$ vs. $\geq 20\%$), the average of nurses' knowledge about risk assessment and effective preventive care per ward (test score <7 vs. ≥ 7), and the main level of nurses' education per ward (intermediate vs. higher vocational training/university). Table 3 shows the scores on every factor per cluster.

Allocating the clusters to a treatment and a control groups using the four methods described above gave the following results. BB resulted in one group containing wards 1, 5, 7, 8, and 10 and a second group containing the other wards with an imbalance of 4. There were 16 other allocation schemes with the same imbalance, but as previously explained, only one (optimal) scheme was randomly selected.

The imbalance was calculated as follows. Under the current allocation scheme, the number of wards with a certain characteristic per group is given in Table 4. The difference is calculated (group A − group B) for every category of every factor. Then, the total imbalance is the sum of the

Table 3. Ward characteristics used as balancing criteria

Ward number	Type[a]	Risk of falling[b]	Test score[c]	Education level[d]
1	1	1	1	1
2	1	0	1	0
3	2	0	1	1
4	2	1	0	0
5	2	1	0	1
6	1	0	1	1
7	1	1	1	0
8	1	0	0	0
9	1	1	0	0
10	2	0	0	0

[a] Type of ward: 1 = surgical, 2 = internal medicine.
[b] Percentage patients with risk of falling: 0 = less than 20%, 1 = 20% or more.
[c] Test of knowledge: 0 = mean score <7, 1 = mean score 7 or higher.
[d] Education level: 0 = intermediate vocational training, 1 = higher vocational training/university.

Table 4. Distribution of prognostic factors of the Safe or Sorry study over treatment groups A and B after BB

Prognostic factor	Group A	Group B	Difference
Type of ward			
Surgical	3	3	0
Internal medicine	2	2	0
Risk of falling (%)			
<20	2	3	−1
≥20	3	2	1
Test score			
<7	2	3	−1
≥7	3	2	1
Education level			
Intermediate vocational training	3	3	0
Higher vocational training/university	2	2	0

Abbreviation: BB, best balance.

squared differences. So, the total imbalance $= 0^2 + 0^2 + (-1)^2 + 1^2 + (-1)^2 + 1^2 + 0^2 + 0^2 = 4$.

Unrestricted randomization resulted in a group containing wards 1, 2, 6, 7, and 9 and the remaining wards in another group, resulting in an imbalance of 52. This imbalance is largely attributable to the fact that all internal medicine wards were assigned to the same group. Furthermore, most of the wards with a low mean test score were also grouped together. So, the distribution of prognostic factors was very dissimilar between the two groups.

Minimization resulted in the same imbalance as BB. With the wards 1, 4, 6, 8, and 10 in one group and the remaining wards in the other, the imbalance was 4.

Next, matching was applied. From the 10,000 matchings that were performed, one of the best matching schemes (including randomization within pairs) was chosen at random. This gave the following pairs: wards 1 and 7, wards 2 and 10, wards 3 and 6, wards 4 and 5, and wards 8 and 9. The randomization within pairs resulted in the same groups of wards as minimization. So, the imbalance of matching was also 4.

Finally, stratification was applied on type of ward and risk of falling because these factors were thought to be most important. Stratifying on a third factor was impossible because this would lead to empty cells or cells containing just one cluster. Randomizing wards within strata resulted in one group containing wards 1, 3, 5, 7, and 8 and a second group containing the remaining wards. This allocation scheme gave an imbalance of 12.

So, in this example, BB, as well as minimization and matching, resulted in an optimal allocation scheme (i.e., one with the smallest imbalance). As mentioned earlier, BB must always result in the smallest imbalance. However, this does not hold for the other methods. All other allocation methods could have ended up with one of the 17 optimal allocation schemes as well, but the probability that this would happen is much smaller for them than that for BB. With unrestricted randomization, for example, this probability is only 0.13 (17 of 126 different allocation schemes).

5. Discussion and conclusion

The aim of this study was to compare BB with unrestricted randomization, minimization, stratification, and matching in terms of achievable balance in CCTs with small numbers of clusters.

A simulation study was performed to compare BB with minimization, randomization, and matching. The results showed that a better balance can be obtained with BB than with the other methods. This effect becomes larger as the number of clusters and number of factors increase.

The results of randomization vs. BB may seem somewhat counterintuitive because randomization is thought to produce treatment groups in which the distributions of prognostic factors are similar, especially when the number of units to randomize is large. However, as the number of clusters increases, so does the number of possible allocation schemes. Therefore, the probability that randomization will result in the best allocation scheme, that is, the one with the best balance of factors, becomes smaller as the number of clusters increases. BB, on the other hand, always searches through all possible allocation schemes. Therefore, the probability of finding the best allocation scheme with this method does not depend on the number of clusters.

We did not evaluate stratification. However, stratification can only take a limited number of factors into account [6]. According to Therneau [6], the maximum number of factors for studies with 4, 8, or 16 clusters is 2, 3, and 4, respectively. The performance of stratification with respect to the obtained balance lies between the performance of unrestricted randomization and minimization, so the balance it achieves will be equal to minimization at best.

Thus, BB is a highly effective method for obtaining balance in prognostic factors over treatment arms. This applies not only to small trials but also to larger trials. However, because the BB method considered here requires all units to be enrolled before allocation, this method is only useful in CCTs where clusters can be identified in advance. If not all, but a substantial part of the clusters is known from the beginning, BB could be used initially. Clusters enrolling later could then be allocated using minimization. This reasoning also applies to studies where individuals are the units of allocation. However, these studies often need large sample sizes. So, it may not be feasible to know all individuals in advance, which makes BB impossible.

The argumentation in this article is based on prognostic factors with only two categories. Nevertheless, the results also apply to factors with more categories. As the number of categories per factor increases, the probability to obtain a good balance will decrease, especially when randomization, stratification, or matching is used. Stratifying on factors with more than two categories will be more difficult and further limits the number of factors that can be taken into account. Matching will also become harder because it will become more difficult to find a good match for all clusters.

BB, on the other hand, is able to take the additional categories into account very well.

The described BB method requires categorization of continuous prognostic factors. Thought should be given to the best limits for the categories because categorization results in a loss of information. Furthermore, the eventual allocation scheme might differ if the limits are chosen differently. If all or most of the factors are categorical from the beginning, the way continuous factors are categorized will be rather unimportant. However, if there are many continuous prognostic factors, minimization based on ranks may be a better method [13,14].

In this study, BB was limited to two treatment arms, but BB can easily be extended to more treatment arms (e.g., if two new treatments are to be compared with a standard treatment). Furthermore, the allocation was limited to a 1:1 ratio. However, the BB algorithm can easily be adapted to other ratios. The key point in each case is to identify the set of all possible allocation schemes and define the imbalance measure, which can be straightforwardly extended from the two-arm 1:1 allocation trial.

In conclusion, BB is a highly effective alternative treatment allocation method in CCTs with small numbers of clusters. Therefore, we advocate the use of this method in such trials.

References

[1] Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in controlled clinical trial. Biometrics 1975;31:103−15.

[2] International conference on harmonisation. Guidance on statistical principals for clinical trials. Fed Regist. International Conference on Harmonisation; E9 Document. 1998;63:49583−98.

[3] Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials. A review. Control Clin Trials 2002;23:662−74.

[4] McEntegart DJ. The pursuit of balance using stratified and dynamic randomization techniques: an overview. Drug Inf J 2003;37:293−308.

[5] Begg CB, Iglewicz B. Treatment allocation procedure for sequential clinical-trials. Biometrics 1980;36:81−90.

[6] Therneau TM. How many stratification factors are too many to use in a randomization plan. Control Clin Trials 1993;14:98−108.

[7] Perry M, Faes M, Reelick MF, Olde Rikkert MG, Borm GF. Study-wise minimization: a treatment allocation method that improves balance among treatment groups and makes allocation unpredictable. J Clin Epidemiol 2010;63:1118−22.

[8] Donner A, Klar N. Design and analysis of cluster randomization trials in health research. London: Arnold; 2000.

[9] Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. BMJ 2004;328:702−8.

[10] D'Agostino RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 1998;17:2265−81.

[11] Taves DR. Minimization—new method of assigning patients to treatment and control groups. Clin Pharmacol Ther 1974;15:443−53.

[12] Borm GF, Hoogendoorn EH, den Heijer M, Zielhuis GA. Sequential balancing: a simple method for treatment allocation in clinical trials. Contemp Clin Trials 2005;26:637−45.

[13] Hoehler FK. Balancing allocation of subjects in biomedical-research—a minimization strategy based on ranks. Comput Biomed Res 1987;20:209−13.

[14] Stigsby B, Taves DR. Rank-minimization for balanced assignment of subjects in clinical trials. Contemp Clin Trials 2010;31:147−50.

[15] van Gaal BG, Schoonhoven L, Hulscher ME, Mintjes JA, Borm GF, Koopmans RT, et al. The design of the SAFE or SORRY? study: a cluster randomised trial on the development and testing of an evidence based inpatient safety program for the prevention of adverse events. BMC Health Serv Res 2009;9:58.