Published in final edited form as:

Contemp Clin Trials. 2013 November ; 36(2): . doi:10.1016/j.cct.2013.09.017.

A Unifying Framework for Standard and Covariate-Adaptive Randomization Procedures Based on Minimizing Suitable Imbalance Functions

Vance W. Berger, PhD

National Cancer Institute and University of Maryland Baltimore County, Biometry Research Group, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20850, (240) 276-7142 (voice)

Vance W. Berger: vb78c@nih.gov

Abstract

Minimization is based on minimizing an imbalance function defined in terms of one or more covariates. Standard (non-adaptive) randomization procedures, on the other hand, generally do not specify or try to minimize an imbalance function. However, it turns out that they may be formulated in this manner. Doing so places adaptive and standard randomization procedures within the same framework, and also suggests novel randomization procedures that combine the best elements of both.

Keywords

Adaptive Randomization; Allocation Concealment; Imbalance Function

1. Introduction

Randomized clinical trials are the gold standard of evidence-based medicine, and the heart of the randomized clinical trial is randomization itself. And yet as important as randomization is, misunderstandings still tend to be the rule, and not the exception. Beyond that, very little attention is paid to the strategic selection of an optimal, or at least a rational, randomization procedure. In fact, the bewildering array of randomization methods used in practice, generally without justification, almost resembles the Tower of Babel, with everyone doing his or her own thing, and speaking his or her own language.

This reality makes it difficult to compare certain randomization procedures to others, which is probably just as well from the perspective of those who use inherently flawed procedures, such as the grossly overly restrictive permuted blocks procedure, which allows for easy prediction of future allocations and, therefore, selection bias [1]. An adaptive procedure such as minimization [2] seems too conceptually different from a non-adaptive procedure, such as the big stick procedure [3], to allow for a meaningful comparison of one to the other. The result is that those who prefer to use minimization use minimization, those who prefer permuted blocks (or, more likely, remain willfully ignorant of the better procedures that are

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

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

readily available) use permuted blocks, those who prefer the big stick use the big stick, and those who prefer the maximal procedure [5] use the maximal procedure. These decisions tend to be made without regard to the merits of the procedure as applied to the specifics of the trial at hand. This lack of uniformity, lack of *adequate* regulation (randomization methods are regulated, of course, but not to the extent that they need to be to ensure that only the best methods are used in practice), and lack of reasoned decision-making results in suboptimal randomization, potentially flawed trials, and distorted results.

One class of non-adaptive randomization procedures is based on a maximally tolerated imbalance (MTI), so that the allocation probabilities are equal when there are equal numbers so far allocated to the two groups and deterministic to the smaller group when the MTI has been met. This class includes the big stick procedure [3], Chen's procedure [4], the maximal procedure [5], and the block urn design [6], all of which offer a solution to the dual problems of 1) ensuring comparable treatment group sizes and 2) minimizing prediction as a way of ensuring allocation concealment and, therefore, minimization of the selection bias that would induce confounding. These dual objectives might be phrased differently as 1) balancing group sizes and 2) balancing covariates. But careful inspection reveals that this second objective actually includes four separate objectives, which might instead be itemized for enhanced clarity. We wish to balance:

1. Group sizes;
2. Key covariates that we know are prognostic;
3. Other measured covariates that may or may not be prognostic;
4. Covariates that the investigator observes but does not measure;
5. Covariates that may not even be observed.

Fayers and Sprangers [7] pointed out that the observed but unmeasured covariates (such as a subjective overall assessment of patient health) can be quite predictive of even hard outcomes, but for our purposes these will be considered interchangeably with those "phantom" covariates that are not observed at all yet still figure so prominently in the detection of selection bias [8], [1]. We shall want to balance all types of covariates, and also treatment group sizes, but we shall not consider the objective of minimizing exposure to that treatment found in this trial to be inferior. The interested reader is referred to Rosenberger and Lachin [9], and references contained therein.

Clearly, minimization is better at balancing a few key covariates than standard randomization procedures are [2], because these standard procedures are omnibus procedures that treat all covariates interchangeably without singling out certain ones as special or as most deserving of being balanced. So if we limit our consideration to only those covariates sufficiently prominent to make it into the imbalance function being minimized, then minimization wins. But minimization has also been rightly criticized [10], [11] on the basis of its precluding the possibility of allocation concealment. In other words, claims by its proponents that its efforts to address the second balancing objective serve also as *de facto* assurances that the third, fourth, and fifth balancing objectives will be addressed are simply not true, and minimization may not win if we consider *all* covariates, even those not in the imbalance function, not measured, and not observed.

We might, at this point, conclude that the different approaches are simply not comparable, and call it a day. If we stopped here, then we could still offer something better than the free-for-all system of anything goes that permeates clinical research currently. Specifically, we would recommend:

1. Stratification if the ratio of the sample size to the product of levels of the key covariates is large enough (and the big stick or maximal procedure within strata);
2. Minimization if we can identify a few key covariates that we know to be jointly sufficiently predictive of the outcome or outcomes, but there are too many levels of these key covariates for stratification to be feasible;
3. An MTI procedure (the big stick procedure or Chen's procedure or the maximal procedure or the block urn design) otherwise.

Under no circumstances should the excessively restrictive permuted blocks procedure be used [12], even if this is the precedent [13], and even if the block sizes are varied [14]. Any of the MTI procedures will be uniformly better, in terms of offering better control of selection bias when matched for the amount of chronological bias allowed [5]. The differences among the MTI procedures are relatively small compared to the benefits they all enjoy over permuted blocks, and are limited to how they handle allocations when there is some imbalance but not enough to reach the MTI condition. In this case, the big stick will allocate equally, Chen's procedure will use a fixed biasing probability that the user specifies, and both the maximal procedure and the block urn design are sensitive to how unbalanced the group sizes are, with more extreme biasing probabilities (closer to 100% for the smaller group) as this level of imbalance increases. The aforementioned decision tree is less than ideal because it does not specify how certain we need to be that the identified covariates are predictive, or how predictive they need to be. Beyond that, this approach also fails to weigh the benefit of balancing these key covariates against the potential harm in the loss of allocation concealment. Nor does it consider the fact that allocation concealment is not ensured even with the big stick or maximal procedures. When we address all of these issues, we find that we can do better.

2. Characterizing Non-Adaptive Procedures in Terms of Imbalance Functions

The permuted blocks procedure with fixed block size two coincides with either the maximal procedure or the big stick procedure if they are used with $MTI=1$ (the MTI is the maximally tolerated imbalance). We may say, then, that the imbalance function is defined in terms of group sizes, namely as the absolute difference between the two group sizes (we shall focus on trials with two treatment groups). Like minimization, then, these procedures randomize only when there is already balance. Otherwise, they allocate deterministically, so as to restore balance. These procedures act very much like minimization, except that they define the imbalance function differently. Hence, as we shall see in the next section, criticism of minimization applies to these procedures too. But the problem is easily resolved by simply increasing the MTI.

Of course, minimization has its extreme form and its less extreme form. The extreme form involves allocating deterministically to restore balance whenever the allocation in question can either increase or decrease the balance. The less extreme form of minimization involves using biasing probabilities to encourage a restoration of balance, without actually forcing it. These biasing probabilities may be fixed, or they may depend on how big an imbalance would result if the allocation were to increase the imbalance. When the MTI exceeds one, the permuted blocks, big stick, and maximal procedures act very much like the less extreme form of minimization, in that they also use allocation probabilities that might be called biasing probabilities. If the MTI is two, then permuted blocks would use blocks of size four, and there are six possible blocks of size four, which are enumerated, along with the conditional allocation probabilities for each allocation of each of the six block types, in Table 1.

If the MTI is three, then there are more block types (of size six), but the point is that if to this point in the block A has been allocated more often than B, then $P\{A\}$ will be less than 0.50, and it falls further away from 0.50 the greater the imbalance. If the imbalance is three, then the MTI has been met and $P\{A\}=0.00$. The specific biasing probabilities do not concern us for the moment. Rather, we focus on the commonality of the permuted blocks procedure and the maximal procedure, which is the monotonicity of $P\{A\}$, dropping from 1.00 when B has been allocated more than A by an amount to match the MTI, $n(B)=n(A)+MTI$, to 0.50 when $n(A)=n(B)$, to 0.00 when $n(A)=n(B)+MTI$. The big stick is contrasted in that all intermediate values between 0.00 and 1.00 are $P\{A\}=0.50$. We see, then, that the permuted blocks and maximal procedures are not so different from the less extreme form of minimization, in that they use $P\{A\}=0.5$ when there is balance, and then biasing probabilities that restore balance when there is imbalance.

These biasing probabilities restore balance more and more forcefully as the imbalance increases, until it reaches the MTI, at which point the allocation is deterministic. Contrast this with the extreme form of minimization, which uses “all or none” allocation probabilities of 0.00 and 1.00 whenever there is any imbalance at all, and with the big stick procedure, which uses balanced allocation probabilities of 0.50 whenever the MTI has not been reached.

3. To What Extent Does Criticism of Minimization Apply More Broadly?

If non-adaptive randomization procedures can be cast in terms of minimizing imbalance functions, then it is a fair question to ask about the extent to which valid criticisms of minimization [10], [11] apply also to non-adaptive procedures. Most of the criticism of minimization focuses on the excessive number of deterministic allocations, owing to the extreme biasing probabilities of 0.00 and 1.00 whenever there is any imbalance at all.

This same criticism applies equally well to any procedure when the MTI is set at one, because then there will be an equal number of deterministic allocations and truly random allocations. But when the MTI increases to a value above one, the maximal and permuted blocks procedures use intermediate biasing probabilities that promote balance without forcing it, whereas the big stick procedure does not even promote balance if there is some imbalance but not quite reaching the MTI. The criticism of minimization, then, is actually a criticism not of the general concept but rather of the specific extreme biasing probabilities. This criticism remains valid when these extreme biasing probabilities are used with procedures other than minimization as well.

4. Hybrid Procedures that Introduce the MTI and Big Stick to Minimization

The imbalance function for minimization need not be integer valued, but to keep the presentation simple, we will consider only imbalances in patient counts by a covariate, for our purposes gender. So suppose that gender is the only known strongly prognostic covariate. We could, if we wanted to, define one imbalance function, $n(A)-n(B)$, for each gender. This would result in stratifying by gender. We might instead define an overall imbalance function of the sort $n(F,A)+n(M,B)-n(F,B)-n(M,A)$, with obvious notation. Now with this imbalance function we can still set an MTI, we can still use deterministic allocations (a big stick) when this MTI is reached, and we can still use equally likely randomization when there is balance. But what do we do when there is some imbalance, but not enough to reach the MTI? Standard minimization, tantamount to an MTI of one, would still use deterministic allocations. The less extreme form of minimization, tantamount to a larger MTI, might follow either Chen [4] and use a constant biasing probability (perhaps 0.75 to the less well represented treatment group) or the maximal (or block urn) paradigm

and use a graded biasing probability depending on how unbalanced the groups are before the allocation.

We can instead define the big stick variant of minimization which uses allocation probabilities of 0.50 no matter what the imbalance (even none at all) as long as the MTI is not reached. We can also define similar permuted blocks and maximal procedure variants of minimization, to try to match the biasing probabilities of these procedures. But it would seem that the big stick variant might be the simplest to plan, to verify, to troubleshoot, to execute, and to explain, so this is the one we shall pursue. The use of this procedure, at least in its pure form, requires the imbalance function to be integer valued. The modification, then, is to introduce an MTI to minimization, and to use deterministic allocations only when this MTI has been reached. Otherwise, we use unpredictable allocations (equal probabilities to each treatment group).

5. A Recommendation on How To Randomize

The key determinants of how one should randomize include:

1. the nature of masking (unmasked, partially masked, or masked to the extent that successful masking is at least possible);
2. the ability or inability to identify key covariates;
3. the number of levels of these key covariates relative to the sample size;
4. the planning of interim analyses or the likelihood of early stopping for any reason;
5. the ease or difficulty of recruiting patients and the expense in treating them according to the protocol.

Consideration of these factors allows for a much more complete decision tree than the preliminary one we presented in Section 1. Before we present this decision tree, we note that the levels of the key covariates may be more parsimonious than they usually are. For example, if there are two key covariates, both binary, say gender and prior surgery, then at first glance this would appear to require four strata ($2 \times 2 = 4$). However, it may turn out that the risk for females with prior surgery is similar to the risk for males without prior surgery, in which case the four strata can safely be reduced to three strata defined by risk group (high, meaning males with prior surgery, low, meaning females without prior surgery, and intermediate, meaning all others).

In general, there will tend to be fewer identifiable risk groups than there are levels of the key covariates, so even if there are too many levels of the key covariates to allow for stratification, stratification may still be an option anyway. For example, if we have three key covariates, all categorical, with three, four, and five levels, respectively, then this would suggest $3 \times 4 \times 5 = 60$ levels, but the key is not to use all $3 \times 4 \times 5 = 60$ levels, but rather to consolidate these into a much smaller number of identifiable risk groups. With this in mind, we shall return to our decision tree.

Specifically, we recommend the following matching of randomization procedures to situations. If the trial is unmasked, or partially masked, but patients are plentiful, and the treatments are not extremely toxic, or expensive, and no interim analyses or early stopping is planned, then it might be a good idea to go with unrestricted randomization, and keep recruiting patients until enough have been allocated to each treatment group. If the trial is unmasked but recruitment is capped by limited resources, then consider the key covariates. If there are very few levels of key covariates, or if a suitably small number of identifiable risk groups can be formed, then stratify, and use a combination [15] of the big stick

procedure and the maximal procedure within strata with an auxiliary randomization to determine which strata use the maximal procedure and which strata use the big stick procedure. These non-adaptive procedures can facilitate the randomization process, since they allow for the allocation sequence to be determined in advance.

Use the big stick variant of minimization if there is a larger number of levels even considering the identification of risk groups, and use the big stick if the key covariates cannot be identified at all. The lack of masking suggests using a large MTI, but this may need to be tempered somewhat if there are plans for early stopping or an interim analysis, because then chronological bias [16] becomes more of an issue. Still, the MTI should not be too small. If the trial is partially masked, then we would use the same decision tree as for unmasked trials, except that now the MTI may be somewhat smaller to better control chronological bias. The MTI might be set based on how successful we expect the masking to be. If we know that the trial will be unmasked, then we want a large MTI, possibly seven or eight. If we strongly suspect that the masking will be successful, then we can get by with a smaller MTI, such as three or two (certainly not one), to minimize chronological bias [16], and relying on the presumed successful masking (which of course must be checked after the fact [8]) to take care of selection bias. If we are unsure, then we would use an intermediate MTI value, perhaps four or five.

6. Conclusions

Randomization is of paramount importance in clinical trials, and this means not only that it be conducted, but also that it be conducted correctly [1], so as to match the situation. One size does not fit all; there is no single randomization procedure that is universally optimal, although there are procedures that are universally flawed, such as permuted blocks, and these should never be used in practice [12], [13], [14]. Even without permuted blocks, there is still a wide array of randomization methods that can be used, and in any given situation, one of these may be better than all others.

We have extended the concept of the MTI, usually restricted to non-adaptive randomization procedures, to minimization, and also worked in the big stick, to introduce the big stick variant of minimization. There is no claim that this will be the best procedure in all contexts, but it may well have its niche, and may be ideal within this niche. We have also proposed a decision tree to infuse reason into the selection process of randomization techniques. There is no claim that the proposed decision tree is optimal, or that it cannot be improved, but it is certainly light years ahead of the anarchy that currently prevails in clinical research, and at the very least would serve as a huge step forward unless or until something better comes along. The days of researchers doing whatever they want, with no justification, must come to an end, hopefully sooner, rather than later, for the sake of the patients ostensibly served by the research. Many parties can, and should, contribute to these needed reforms. One step will be improved randomization methods in all randomized trials, and this may come about at the initiative of the researchers themselves or because the need for increased rigor is imposed on them externally by funding agencies and/or regulators and/or journal editors.

Acknowledgments

I thank the review team for helpful comments.

References

1. Berger, VW. Selection Bias and Covariate Imbalances in Randomized Clinical Trials. Chichester: John Wiley & Sons; 2005.

2. Scott NW, McPherson CG, Ramsay CR, Campbell MK. The Method of Minimization for Allocation to Clinical Trials: A Review. *Controlled Clinical Trials*. 2002; 23(6):662–674. [PubMed: 12505244]
3. Soares JF, Wu CFJ. Some Restricted Randomization Rules in Sequential Designs. *Communications in Statistics Theory and Methods*. 1983; 12:2017–2034.
4. Chen YP. Biased coin design with imbalance tolerance. *Communications in Statistics*. 1999; 15:953–975.
5. Berger VW, Ivanova A, Deloria-Knoll M. Minimizing Predictability while Retaining Balance through the Use of Less Restrictive Randomization Procedures. *Statistics in Medicine*. 2003; 22(19):3017–3028. [PubMed: 12973784]
6. Zhao W, Weng Y. Block urn design - a new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. *Contemp Clin Trials*. 2011 Nov; 32(6): 953–61. [PubMed: 21893215]
7. Fayers PM, Sprangers MA. Understanding Self-Rated Health. *Lancet*. 2002; 359:187–188. [PubMed: 11812551]
8. Berger VW, Exner DV. Detecting Selection Bias in Randomized Clinical Trials. *Controlled Clinical Trials*. 1999; 20:319–327. [PubMed: 10440559]
9. Rosenberger, W.; Lachin, JM. *Randomization in Clinical Trials: Theory and Practice*. New York: John Wiley and Sons; 2002.
10. Berger VW. Minimization, by its Nature, Precludes Allocation Concealment and Invites Selection Bias. *Contemporary Clinical Trials*. 2010; 31:406. [PubMed: 20457277]
11. Berger VW. Minimization: Not All It's Cracked Up To Be. *Clinical Trials*. 2011; 8:443. [PubMed: 21835864]
12. Berger VW. Do Not Use Blocked Randomization. *Headache*. 2006; 46:2, 343.
13. Berger VW. Misguided Precedent Is not a Reason To Use Permuted Blocks. *Headache*. 2006; 46(7):1210–1212. [PubMed: 16866731]
14. Berger VW. Varying Block Sizes Does Not Conceal the Allocation. *Journal of Critical Care*. 2006; 21(2):229. [PubMed: 16769475]
15. Berger VW, Grant WC, Vazquez LF. Sensitivity Designs for Preventing Bias Replication in Randomized Clinical Trials. *Statistical Methods in Medical Research*. 2010; 19(4):415–424. [PubMed: 20488837]
16. Matts JP, McHugh RB. Conditional Markov Chain Designs for Accrual Clinical Trials. *Biometrical Journal*. 1983; 25:563–577.

Table 1

Biasing Probabilities for Permuted Blocks of Size Four

Sequence	Biasing Probabilities $P\{A\}$
AABB	0.50, 0.33, 0.00, 0.00
ABAB	0.50, 0.33, 0.50, 0.00
ABBA	0.50, 0.33, 0.50, 1.00
BAAB	0.50, 0.67, 0.50, 0.00
BABA	0.50, 0.67, 0.50, 1.00
BBAA	0.50, 0.67, 1.00, 1.00