Contemporary Clinical Trials Communications 2 (2016) 80–84

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Characterizing permuted block randomization as a big stick procedure

Vance W. Berger [a, *], Isoken Odia [b]

[a] National Cancer Institute and University of Maryland Baltimore County, Biometry Research Group, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20850, USA
[b] University of South Florida, 13301 Bruce B. Downs Boulevard, Tampa, FL 33612, USA

ARTICLE INFO

Article history:
Received 7 September 2015
Received in revised form
14 December 2015
Accepted 15 January 2016
Available online 29 January 2016

Keywords:
Big stick
Blocked randomization
Maximal procedure

ABSTRACT

There are numerous approaches to randomizing patients to treatment groups in clinical trials. The most popular is permuted block randomization, and a newer and better class, which is gaining in popularity, is the so-called class of MTI procedures, which use a big stick to force the allocation sequence back towards balance when it reaches the MTI (maximally tolerated imbalance). Three prominent members of this class are the aptly named big stick procedure, Chen's procedure, and the maximal procedure. As we shall establish in this article, blocked randomization, though not typically cast as an MTI procedure, does in fact use the big stick as well. We shall argue that its weaknesses, which are well known, arise precisely from its improper use, bordering on outright abuse, of this big stick. Just as rocket powered golf clubs add power to a golf swing, so too does the big stick used by blocked randomization hit with too much power. In addition, the big stick is invoked when it need not be, thereby resulting in the excessive prediction for which permuted blocks are legendary. We bridge the gap between the MTI procedures and block randomization by identifying a new randomization procedure intermediate between the two, namely based on an excessively powerful big stick, but one that is used only when needed. We shall then argue that the MTI procedures are all superior to this intermediate procedure by virtue of using a restrained big stick, and that this intermediate procedure is superior to block randomization by virtue of restraint in when the big stick is invoked. The transitivity property then completes our argument.

1. Introduction

There are numerous approaches to randomizing patients to treatment groups in clinical trials, the most popular of these being permuted block randomization, which holds a near monopoly on how trials are randomized in practice [1–5]. For example, randomization.com, when used in this context, offers only what it calls "randomly permuted blocks" and no other alternatives [6]. Even so authoritative a document as the International Conference on Harmonization Guideline E9 (Statistical Principles for Clinical Trials, February 5, 1998) [7] recommends randomizing subjects in blocks. Given this popularity and near ubiquity of permuted block randomization, one might expect, on this basis alone, that the permuted blocks design is also, in some sense, optimal, since it is fair to ask why it would be used so often if this was not the case. Unfortunately, this is actually not the case, and for this reason many competitors have been proposed [2], [3], [8–10]. One newer and better class, which is gaining in popularity, is the so-called class of MTI procedures, which use a big stick to force the allocation sequence back towards balance when it reaches the MTI (maximally tolerated imbalance). Four prominent members of this class are the aptly named big stick procedure [11], Chen's procedure [12], the maximal procedure [13], [14], and the block urn design [3] which shares many desirable properties with the maximal procedure.

As we shall establish in this article, permuted block randomization, though not typically cast as an MTI procedure, does in fact use the big stick as well. We shall argue that its well-known inability to ensure comparable comparison groups arises precisely from its improper use, bordering on outright abuse, of this big stick. Just as rocket powered golf clubs add power to a golf swing, so too does the big stick used by permuted block randomization hit with too much power. In addition, the big stick is invoked when it need not (and *should* not) be, thereby resulting in the excessive prediction, selection bias (which can arise even in the absence of any malice or intention to bias the trial), and baseline imbalances for which permuted blocks are legendary [4], [5]. These

* Corresponding author.
E-mail addresses: vb78c@nih.gov (V.W. Berger), iodia@health.usf.edu (I. Odia).

http://dx.doi.org/10.1016/j.conctc.2016.01.001

are two distinct abuses of the big stick that lead to excess vulnerability to prediction.

Consideration of the two weaknesses together may tend to muddy the waters, so we try here to bring clarity by isolating each one, and considering it on its own merits. This allows us to bridge the gap between the MTI procedures and block randomization by identifying a new randomization procedure that serves as the missing link, intermediate between the two, namely based on an excessively powerful big stick (alluded to in Ref. [15]), but one that is used only when needed. We shall then argue that the MTI procedures are all superior to this intermediate procedure by virtue of using a restrained big stick, and that this intermediate procedure is superior to block randomization by virtue of restraint in when the big stick is invoked. The transitivity property then completes our argument.

2. Rocket powered big sticks

Let us suppose that we can agree on a boundary. A chess board has 64 squares, so a piece may not legally move any further than that. The dimensions of a soccer field are marked off in advance of the game, so that all parties can agree on where the line is drawn, so to speak. These are examples of reflective boundaries. If a bishop moves to the perimeter (outside) of the chess board, then the next time it moves, that move has to be away from that boundary, not over it. Likewise, barring a red card, when a soccer player dribbles the ball to the edge of the soccer field, his or her next move (with respect to that boundary; we consider movement up and down the line to be no movement at all with respect to that line, as in projecting a two-dimensional space onto one dimension) must be away from that edge, and not over it. This much is clear, and there is also a rather clear parallel with allocation sequences when viewed as random walks.

The question before us is how far away from the boundary (as defined by the MTI condition) must one be sent upon reaching it? The rules of chess, for example, could be modified so that any time a king, queen, rook, bishop, or knight reaches any edge of the board (with the possible exception of the starting position), it must travel from there directly towards one of the central four squares, much the same way that the reflecting barrier works in monopoly, with the "go to jail" property serving as the boundary and the jail itself serving as center. In monopoly, the piece does not simply move back one property; it gets sent all the way back.

Nor is monopoly the only such example; in black jack, get to 21 and keep going and you get sent all the way back to zero (busted, as some call it). In soccer, the rules could state that whenever the ball reaches the edge of the field, it must be taken from there to the center circle. That would, of course, dampen the excitement surrounding corner kicks. Chess, soccer, black jack, and monopoly are just examples. We cannot on that basis make a determination regarding what is and is not appropriate for randomization in clinical trials, as the situations are entirely different. The games are presented as parallels only to illustrate the distinction between forced returns *to* the center and forced returns *towards* the center. Strictly on its own merits, which one makes more sense in our context, randomization in clinical trials? The standard big stick, to knock the allocation sequence back towards balance *by one unit*, or the super charged big stick that blasts the sequence all the way back towards perfect balance each time it dares to reach the edge? We cannot answer this question in a vacuum. Rather, we must consider the purpose of the reflecting boundary. Why is it so important that the boundary not be crossed? And is there commensurate harm caused by mere proximity to the boundary? Or is it instead a threshold effect, kicking in only when the boundary is actually crossed?

The key idea here is chronological bias [16], which has been adequately described in the literature. Briefly, we do not wish to allow the numbers of patients allocated to each treatment group to differ by too much at any one point in time, because this, coupled with time trends in key indicators of disease severity, can result in a substantial baseline imbalance across treatment groups, or confounding. Operationally, we deal with this by specifying a maximum tolerated imbalance (MTI). This is the reflecting boundary.

For example, even though one does not generally speak of an MTI when permuted blocks are used, blocks of size four (with two treatment groups) will induce an MTI of two. In general, when permuted blocks are used with any fixed block size (and two treatment groups), the MTI is half that block size. When varied block sizes are used, the MTI is half the largest block size (with two treatment groups), or the largest block size divided by the number of treatment groups (assuming equal allocation). In full generality, although this rarely comes up, the MTI induced by the use of permuted blocks, with arbitrary number of treatment groups and arbitrary allocation ratios, is the product of the largest block size and the largest allocation ratio for any one treatment group. So, for example, with three treatment groups, and varied block sizes of five and ten, with allocations in the ratio of 2:2:1, the MTI is (10) (2/5) = 4, since this randomization plan exposes us to the risk of an initial sequence of four consecutive allocations to A.

It is objectively clear that large imbalances at any point in time can lead to problems, as already discussed. It is less clear where to draw the line, since this is not really a binary phenomenon. If we use an MTI of four, then is this to suggest that an imbalance of three (or even four itself) is *not* a problem? What if we keep hitting the MTI boundary again and again, so we reach this level of imbalance fairly often during the course of the patient allocation process? One could certainly argue that there is harm done not only by crossing the MTI threshold but also by pressing right up against it repeatedly, especially if these swings are always in the same direction.

That is to say that if the MTI is three, then AAABBBAAABBBAAABBB might be considered a problematic allocation sequence, even while not crossing the boundary and therefore not being disallowed, since the accession numbers associated with treatment group A are systematically smaller than those associated with treatment group B. In other words, there are systematically more early allocations to treatment A and fewer to treatment B, so therefore this allocation sequence might also be considered worse than AAABBBBBBAAAAAABBB, which hits the boundary equally often, but hits both sides of the MTI boundary, rather than hitting the same side repeatedly, so that imbalances go in both directions over the course of the trial, rather than always going in the same direction. There would be an equal number of overall imbalances, but these are at least spread more evenly across the two treatment groups.

But is this *enough* of a problem to merit the draconian measures used to curtail it? If these sequences (and others like them) could be avoided with no dire consequences, then we would be in favor of eliminating them. But there *is* a cost, and a rather steep one at that. Each forced allocation is a deterministic allocation, and these are predictable, and have the potential to lead to selection bias by eliminating the possibility of allocation concealment [10]. It is not possible to simultaneously eliminate both chronological bias and selection bias [9]. So which one represents the more serious threat to the integrity of the trial? It seems fairly clear that selection bias is the more serious issue, since it can be steered in a preferred direction by a zealous investigator. That is, it is a true bias. But chronological bias, despite its name (a misnomer, actually), is *not* a bias. It is equally likely to go either way, and it is hard to imagine any plausible scenario under which an investigator exploits it to