Creating Comparable Comparison Groups in Randomized Clinical Trials

Vance W. Berger, NCI

10/25/16, BASS, Rockville

Abstract

- Randomization is one of the cornerstones of the randomized clinical trial, and there is no shortage of methods one can use to randomize patients to treatment groups. When deciding which one to use, researchers must bear in mind that not all randomization procedures are equally adept at achieving the objective of randomization, namely balanced treatment groups. One threat is chronological bias, and permuted blocks randomization does such a good job at controlling chronological bias that it has become the standard randomization procedure in clinical trials. But permuted blocks randomization is especially vulnerable to selection bias, so as a result various MTI procedures have been proposed as better alternatives. These include the big stick, Chen, maximal, and, most recently, asymptotic maximal procedures. In comparing them, we have somewhat of a false controversy, in that actual practice goes uniformly one way (permuted blocks), whereas scientific arguments go uniformly the other way (MTI procedures). There is no argument in the literature to suggest that the permuted block design is better than or even as good as the MTI procedures, but this dearth is matched by an equivalent one regarding actual trials using the MTI procedures. So the "controversy", if we are to call it that, pits misguided precedent against sound advice that tends to be ignored in practice. We shall review the issues to determine scientifically which of the two approaches is better and, therefore, should be used.

Outline

- 1. Restricted Randomization (2).
- 2. Permuted Blocks (3).
- 3. Selection Bias Mechanism (5).
- 4. MTI Randomization Procedures (3)
- 5. Encryption in All its Forms (4).
- 6. The Maximal Procedure (2).
- 7. The Asymptotic Maximal Procedure (5).
- 8. Summary (3)
- 9. Further reading (3).

1. Restricted Randomization (1/2)

- Clinical trials ideally compare treatment groups that are comparable at baseline.
- But randomization alone does not always create the balanced groups we hope for.
- Because of baseline imbalances, the ***success of randomization*** has been questioned in many RCTs [1], including The University Group Diabetes Program [2] and mustine v talc for control of pleural effusions [3], [4].

1. Restricted Randomization (2/2)

- Completely unrestricted randomization allows for unbalanced group sizes, and so is not used very often.
- Instead, some form of restricted randomization is used to ensure balanced group sizes at the end of the trial.
- The random allocation rule forces only terminal balance in group sizes, so it allows for large baseline imbalances *during* the trial, and, therefore, chronological bias [5].
- Suppose that many more early allocations are to one group, and more late allocations are to the other group.
- Suppose further that the covariate distribution changes during the course of the trial; this is quite likely.
- The only way to control chronological bias is to introduce *more* restrictions on the randomization.

2. Permuted Blocks (1/3)

- Permuted blocks stratify by order of arrival.
- If the block size is four, e.g., then the first four patients enrolled (in each stratum) constitute a block, and two patients per block are allocated to each group to force perfect balance in each block.
- With block size two and treatments E and C, there are two block types, CE and EC, with P{E}={0.5, 1.0} for CE and P{E}={0.5, 0.0} for EC.
- Randomized blocks minimize chronological bias by ensuring that the imbalance is bounded.

2. Permuted Blocks (2/3)

- But in unmasked trials, prior allocations are known; this is a concern even in imperfectly masked trials (and how often is masking perfect?).
- Once all but one group has been exhausted in the block (e.g., EE***CC*** with size 4), all remaining allocations to that block will be deterministic.
- In an EECC block even the 2nd is *predictable*, as there is a 2/3 chance of this allocation being C.
- The problem is that the upcoming treatment is known *before* the patient is identified to fill it.

2. Permuted Blocks (3/3)

- Let r(n)=r(n;E)+r(n;C) be the remaining number of allocations in the block just prior to the n^{th} allocation, and define P{E}=r(n;E)/r(n).
- Clearly, this conditional allocation probability is not always equal to the unconditional one.
- With 1:1 allocation and block size 4 P{E} is:
- CCEE 2/4, 2/3, 2/2, 1/1 EECC 2/4, 1/3, 0/2, 0/1
- CECE 2/4, 2/3, ½. 1/1 ECEC 2/4, 1/3, ½, 0/1
- CEEC 2/4, 2/3, ½, 0/1 ECCE 2/4, 1/3, ½, 1/1
- There is quite a bit of prediction going on, and this is true *even when allocation concealment is (incorrectly) claimed*.

3. Selection Bias Mechanism (1/5)

- The process of randomization is nothing more, or less, than constructing treatment groups by randomly selecting non-overlapping subsets of the set of all accession numbers to be used [6].
- This is only the first step for ensuring balance.
- The second line of defense is often masking.
- The third line of defense is superficial or prima facie allocation concealment, which is often defined as masking each allocation only until a treatment is assigned to the patient in question.
- This step is almost universal now, so we must focus more on prediction than on observation.

3. Selection Bias Mechanism (2/5)

- The process of masking, or not telling patients or physicians who got what, is clearly worthwhile, but information may not be contained very well.
- Tell-tale side effects, e.g., may lead to unmasking.
- Sealed envelopes have been held up to lights, files have been raided, and fake patients have been called in to ascertain the next allocation [6].
- So the effect of masking may not match its goal.
- Most RCTs use restricted randomization (blocks).
- The patterns in the allocation sequence allow for selection bias through prediction of the future allocations based on knowledge of past ones [6].

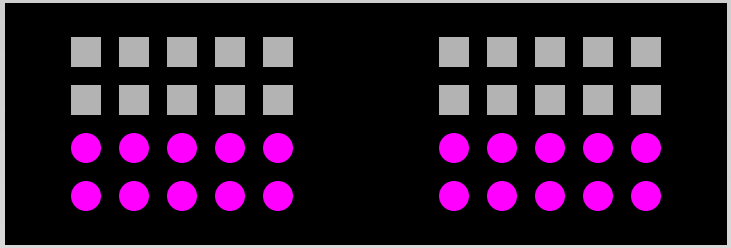
3. Selection Bias Mechanism (3/5)

- One can compute the expected imbalance in a binary covariate to be 50% with blocks of size 2, 42% (block size 4), or 28% (block size 6) [7].
- We then get artificially large test statistics and posterior probabilities, artificially low p-values, and artificially narrow confidence intervals.
- Note that this may well be the case with *all* endpoints, and *all* measures, so the net result is a false sense of security from consistency of the results (all of which are distorted by the bias).
- Even randomized trials with prima facie allocation concealment are not immune [6]; see the figure below, which depicts 10 EC blocks, 10 CE blocks.

3. Selection Bias Mechanism (4/5)

All Patients To Be Randomized (20 Male, 20 Female, 10 CE Blocks, 10 EC Blocks))

20 blocks of size two each
10 ‘CE’ blocks, 10 ‘EC’ blocks
For ‘CE’, P{E}={0.5, 1.0}
For ‘EC’, P{E}={0.5, 0.0}
Females respond better than males



Selectively Semi-permeable

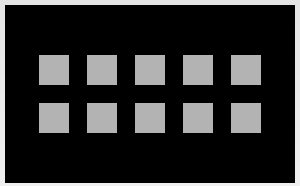
Permeable

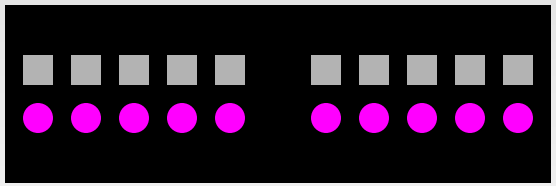
Selectively Semi-permeable

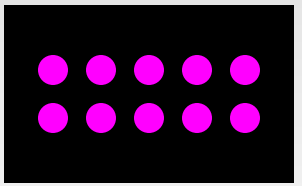
P{E}=0.0 (10 Male)

P{E}=0.5 (10 Male, 10 Female)

P{E}=1.0 (10 Female)





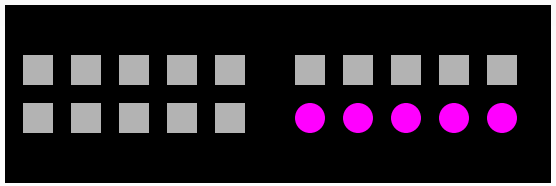


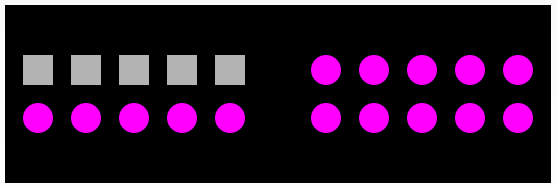
100%

50%

50%

100%





Control Group
(25% Female, 75% Male)

Experimental Group
(75% Female, 25% Male)

3. Selection Bias Mechanism (5/5)

- Recall the P{E}=0.5 for the first patient in either a CE or an EC block, and P{E}=0 or 1 after that.
- We depicted gender as the selection variable for simplicity, but in practice this would probably be too obvious, since gender is always checked.
- But subjective health perceived by a patient can predict clinical outcomes and even mortality, even after adjusting for other observed predictors [8].
- Instead of gender, consider an overall assessment (not committed to the CRF or data base).

4. MTI Procedures (1/3)

- In light of the susceptibility of permuted block randomization to selection bias, a new class of randomization procedures has been developed.
- In fact, various authors have discovered this superior approach independently of each other.
- MTI randomization is based on a maximally tolerated imbalance (MTI) specified by the user.
- With two treatment groups, allocation is 50:50 when there is balance, and deterministic (100% to the smaller group) when the MTI is reached; the procedures differ only with lesser imbalances.

4. MTI Procedures (2/3)

Randomization Procedure (MTI=3)						
Imbalance	Big Stick		Chen (refinement of the Big Stick)		Maximal (refines Chen)	
	Smaller Group	Larger Group	Smaller Group	Larger Group	Smaller Group	Larger Group
0	50%	50%	50%	50%	50%	50%
1	50%	50%	P>50%	1-P	P1>50%	1-P1
2	50%	50%	P>50%	1-P	P2>P1	1-P2
3	100%	0%	100%	0%	100%	0%

4. MTI Procedures (3/3)

- To illustrate the differences among the MTI procedures, consider a trial with n=4 and MTI=2.
- There are then 12 admissible sequences, of which we list six, since the other six are mirror images.
- EECE, EECC, ECEE, ECEC, ECCE, ECCC.
- The maximal procedure [9] picks one of the 12 (mirror images too), with equal probabilities, 1/12.
- If the first allocation is E, then P{C}=4/6 for the second allocation with the maximal procedure.
- The big stick [10] uses P{C}=P{E}=0.5.
- Chen's procedure [11] uses a specified value p.

5. Encryption (1/4)

- Consider a military analogy, encrypted codes.
- Obviously, no army is going to hand over a secret battle plan to the enemy; so no need to encrypt?
- Robert E Lee did not expect Special Order #191 to fall into the wrong hands, so he did not code it.
- How differently might things have turned out?
- The military began to recognize the need for a second level of defense, such as the Enigma.
- Consider the allocation sequence the battle plan.
- It will not be displayed, but it can still be deduced.

5. Encryption (2/4)

- The question is how much can one auto-predict?
- MTI randomization offers superior **overall encryption** than the permuted blocks procedure.
- There will be far fewer deterministic allocations.
- But the benefits do not end there; we must also consider **predictable encryption**, as in, with MTI randomization one cannot determine in advance which allocations can be predicted.
- Compare this to blocks, which are deterministic at regular intervals that are known in advance.

5. Encryption (3/4)

- The sharpest contrast here occurs if we consider a trial that retains the masking for the first few patients, but then not for the subsequent ones.
- Let us say that the MTI is two in this case, and only the first eight allocations remain masked.
- In this case, with blocks, one can easily pick up the prediction starting with the third block.
- But with an MTI procedure, one is completely lost, and it will be nearly impossible to predict.
- MTI procedures enjoy **elastic encryption**; any masking will reduce prediction; not so for blocks.

5. Encryption (4/4)

- In addition to the inherent encryption benefits of the MTI procedures, there is also familiarity.
- Everyone is familiar with blocks, and assumes that they will be used in all trials; prediction starts with the first patient randomized (no learning curve).
- MTI procedures offer **familiarity encryption**.
- Even without masking, thwart enough early attempts at prediction and there may be no more.
- Combining 1) an MTI procedure; 2) concealing the MTI; and 3) even only partial masking can completely *eliminate* selection bias.

6. The Maximal Procedure (1/2)

- Unlike the big stick procedure, the maximal procedure selects at random, with equal odds, from the set of admissible allocation sequences.
- This induces the biasing probabilities, which are *not* user specified, as they are with Chen.
- So balance is encouraged (even if not forced) when there is any imbalance, even if it does not reach the MTI, and, unlike Chen, is encouraged more forcefully as the imbalance increases [9].
- If MTI=2, E need not follow ECC; if MTI=3, C need not follow EEEC (compare to blocks).

6. The Maximal Procedure (2/2)

- Without any formal analysis, and using only intuition, you might already suspect that Chen is better than the big stick by virtue of refining it.
- Likewise, the maximal is intuitively more appealing than Chen, again, as a refinement.
- But do we need the maximal procedure, or is any other refinement (with proper monotonicity) just as good, as long as P2>P1>50%?
- Note that the maximal biasing probabilities are unstable or non-Markovian, meaning that they vary over the course of the trial [12].

7. The AM Procedure (1/5)

- In addition, even just changing the sample size of the trial will also alter the maximal probabilities.
- There are reasons for this, and it can be explained, and can even be cast as a desirable feature.
- However, this feature also renders the maximal procedure (and block urn [13]) difficult to automate, and to explain to clinical colleagues.
- One saving grace is that the maximal biasing probabilities do converge as the trial gets larger.
- The asymptotic maximal procedure [12] uses these limiting biasing probabilities throughout the trial.

7. The AM Procedure (2/5)

Comparison between Randomization Designs for Two Arm Equal Allocation Trials with Maxmial Tolerated Imbalance of **2**						
Current Imbalance	Conditional Allocation Probability Pr(T=A)					
N_A-N_B	Permuetd Block (Block Size = 4)	Big Stick	Biased Coin (P_{bc} = 0.667)	Biased Coin (P_{bc} = 0.8)	Block Urn Design (Block Size = 4)	Asymptotic Maximal Procedure
-2	**1**	**1**	**1**	**1**	**1**	**1**
-1	0.667; **1**	0.5	0.667	0.8	0.667	0.667
0	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
1	0.333; **0**	0.5	0.333	0.2	0.333	0.333
2	0	0	0	0	0	0
Proportion of Deterministic Assignment	33.3%	25%	16.5%	9.9%	16.7%	16.7%
Correct Guess Probability	70.8%	62.5%	66.6%	70.0%	66.7%	66.7%

7. The AM Procedure (3/5)

Comparison between Randomization Designs for Two Arm Equal Allocation Trials with Maxmial Tolerated Imbalance of **3**						
Current Imbalance	Conditional Allocation Probability Pr(T=A)					
N_A-N_B	Permuetd Block (Block Size = 6)	Big Stick	Biased Coin (P_{bc} = 0.667)	Biased Coin (P_{bc} = 0.8)	Block Urn Design (Block Size = 6)	Asymptotic Maximal Procedure
-3	**1**	**1**	**1**	**1**	**1**	**1**
-2	0.75; **1**	0.5	0.667	0.8	0.75	0.707
-1	0.6; 0.667, **1**	0.5	0.667	0.8	0.6	0.586
0	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
1	0.4; 0.333; **0**	0.5	0.333	0.2	0.4	0.414
2	0.25; **0**	0.5	0.333	0.2	0.25	0.293
3	0	0	0	0	0	0
Proportion of Deterministic Assignment	25.0%	16.70%	7.1%	2.30%	5.9%	7.2%
Correct Guess Probability	68.3%	58.3%	64.2%	69.0%	63.2%	62.3%

7. The AM Procedure (4/5)

7. The AM Procedure (5/5)

- We see that, no matter which randomization method is used, larger MTI values work to reduce vulnerability to selection bias.
- We also see that blocked randomization shares the undesirable feature of unstable allocation probabilities with the maximal procedure.
- Blocked randomization is far worse than the others, even when blocks sizes are varied.
- Chen with P=0.99 is ideal to prevent deterministic allocations, the big stick is best against predictable allocations, and the AM is best overall [12].

8. Summary (1/3)

- The creation of comparable comparison groups is arguably the key element of trial quality.
- Permuted block randomization simply cannot assure comparable comparison groups.
- It is way too vulnerable to prediction even with large and/or varied block sizes; nor does a large sample size help in any measureable way.
- Any MTI procedure will be much better.
- The use of blocks is no longer defensible, and this remains true even in ostensibly masked trials, and even with large and/or varied block sizes.

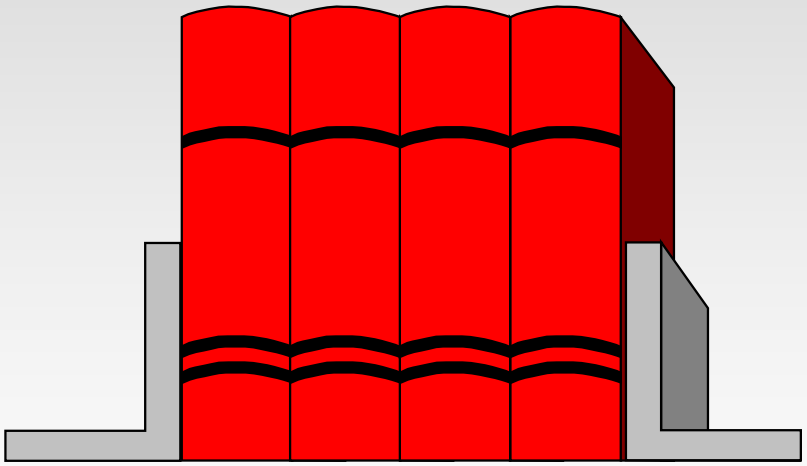
8. Summary (2/3)

- Unless or until a better option is found, an MTI randomization procedure should be used.
- But among the MTI procedures, there is no clear cut optimal one, as each one has its own merits.
- Any one of them is defensible; I would not object!
- Alas, we must choose one; the asymptotic maximal procedure appears to be the most robust.
- The procedure itself is easily understood, easy to explain, and now also easy to use in practice [14].
- At this point I would like to ask for objections to using the maximal procedure routinely for trials.

8. Summary (3/3)

- You CAN make a huge difference in helping to improve the deplorable situation in which just about all trials randomize inappropriately.
- As a reviewer, remain resolute that there is no place in serious research for demonstrably fatally flawed methods such as blocked randomization.
- As a researcher, use MTI randomization only.
- The Berger-Exner test of selection bias should also be used routinely after the trial is over.
- Contact me about collaboration and to let me know of any trials with suspected selection bias.

9. Further Reading (1/3)



- More information is available -- just send me a message and I will send you articles.

- Vance Berger
- Vb78c@nih.gov
- (301) 435-5303

9. Further Reading (2/3)

- [1]. Berger VW, Weinstein S (2004). Ensuring the Comparability of Comparison Groups: Is Randomization Enough? *Controlled Clinical Trials* **25**, 515-524.
- [2]. Schor, S. (1971). The University Group Diabetes Program: A Statistician Looks at the Mortality Results. *JAMA* **217**, 12, 1671-1675.
- [3]. Altman, D. G. (1985). Comparability of Randomized Groups. *The Statistician* **34**, 125-136.
- [4]. Fentiman, I. S., Rubens, R. D., Hayward, J. L. (1983). Control of Pleural Effusions in Patients with Breast Cancer. *Cancer* **52**, 737-739.
- [5]. Matts, J. P. and McHugh, R. B. (1983). Conditional Markov chain design for accrual clinical trials. *Biometrical Journal* **25**, 563-577.
- [6]. Berger, VW (2005). *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*, John Wiley & Sons, Chichester.
- [7]. Berger, VW (2005). "Quantifying the Magnitude of Baseline Covariate Imbalances Resulting from Selection Bias in Randomized Clinical Trials" (with discussion), *Biometrical Journal* **47**, 2, 119-139.

9. Further Reading (3/3)

- [8]. Fayers PM, Sprangers MA. Understanding self-rated health. *Lancet* 2002; 359:187-188.
- [9]. Berger, VW, Ivanova, A, Deloria-Knoll, M (2003). “Minimizing Predictability while Retaining Balance through the Use of Less Restrictive Randomization Procedures”, *Statistics in Medicine* **22**, 19, 3017-3028.
- [10]. Soares, JF, Wu, CFL (1983). “Some Restricted Randomization Rules in Sequential Designs”, *Communications in Statistics Theory and Methods* **12**, 2017-2034.
- [11]. Chen, YP (1999). “Biased Coin Design with Imbalance Intolerance”, *Communications in Statistics Stochastic Models* **15**, 953-975.
- [12]. Zhao, WL, Berger, VW, Yu, Z (2017). "The Asymptotic Maximal Procedure for Subject Randomization in Clinical Trials", *Statistical Methods in Medical Research*, in press.
- [13]. Zhao W, Weng Y. Block urn design – A new randomization algorithm for sequential trilas with two or more treatments and balanced or unbalanced allocation. *Contemporary Clinical Trials* 2011; **32**:953-61.
- [14]. https://cran.r-project.org/web/packages/randomizeR/

3. Selection Bias Mechanism (2/6): Allocations Revealed