CLINICAL INVESTIGATION

A note on conditional and unconditional allocation concealment

Vance W Berger*
*National Cancer Institute & University of Maryland Baltimore County Biometry Research Group, National Cancer Institute 9609 Medical Center Drive Rockville, MD 20850, USA
Tel: +1 240 276 7142 vb78c@nih.gov

Allocation concealment is often misunderstood but, when understood properly, can offer a metric for evaluating the quality of nonadaptive randomization procedures. For example, appeal to allocation concealment tells us that the maximal procedure and the big stick procedure are both superior to permuted blocks. However, the standard definition of allocation concealment is not sufficient to properly evaluate adaptive randomization, be it covariate adaptive or response adaptive randomization. We therefore propose a new concept, namely conditional allocation concealment, to be distinguished from conventional or unconditional allocation concealment, to better understand the strengths and weaknesses of existing adaptive randomization methods. Better evaluation of randomization methods will predictably lead to better randomization, too, and this will in turn lead to more robust trials.

Keywords: adaptive randomization • allocation concealment • minimization

A widespread misunderstanding of allocation concealment contributes to its improper use in practice, and makes it difficult to propose new advances, as one must then decide on whether to try to improve the state of the art that is essentially never used (and rarely understood), or the basic level of understanding that permeates clinical trials research. The common understanding of allocation concealment is that it is violated when, and only when, the upcoming allocations can be directly observed, as was the case in the Canadian National Breast Cancer Screening Study [1]. This would mean that allocation concealment is ensured any time the allocation list is concealed and, therefore, that all it takes to make the claim of allocation concealment is this effort.

Unfortunately the reality is more complicated, and there are two threats, not only one threat, to allocation concealment [2,3,4]. Specifically, there is no allocation concealment if upcoming allocations can be predicted, even if they are never directly observed. This prediction is rooted in the specific manner in which the randomization is conducted. For example, in an unmasked trial using permuted blocks of size two, every second allocation can be predicted with certainty even if the standard understanding of allocation concealment is satisfied. That is, even if the allocation sequence is kept hidden, the use of permuted blocks still precludes the possibility of allocation concealment in an unmasked or imperfectly masked trial.

In this regard, it would be better to increase the block size, because then fewer allocations would be predictable. Still, some would be, and we see that allocation concealment can be elusive, is not established with just the claim and is not a binary phenomenon. It can be violated to a greater or lesser extent depending on how much unmasking there is, and also on the specific restrictions used in the preparation of the allocation sequence. We can hold one of these variables constant and vary the other one, to obtain useful results. Specifically, we ask which randomization procedures are better than others for minimizing prediction in a completely unmasked trial.

This question has already been asked and answered [5,6,7], and the conclusion was that unrestricted randomization is best for this

FUTURE SCIENCE part of fsg

purpose, although this may not be considered practical if comparable group sizes are a requirement. So the idea is to use the fewest restrictions possible subject to forcing comparable group sizes, and this means that the maximal [5] and big stick [7] procedures are best, and the permuted blocks procedure has no place in serious clinical research. But this formulation does not really address adaptive procedures in general, or minimization in particular.

Does minimization allow for allocation concealment

Minimization differs from the aforementioned randomization procedures in that these other procedures treat all covariates interchangeably and attempt to balance them, whereas minimization singles out certain covariates as more important than others and then proceeds to force the balance of these chosen covariates. This may be done in a deterministic fashion (except for when the same imbalance results no matter how the next patient is allocated) or in a random fashion with biasing probabilities that are not quite as extreme as zero and one. These biasing probabilities may or may not reflect the difference in the level of imbalance that would result in randomization of the next patient to either of the two treatment groups.

When using minimization, or basically any other type of adaptive randomization, one can make the argument that the investigators have no idea which treatment will be allocated next, and, therefore, that we are operating with complete allocation concealment. And yet this statement would not do justice to the legitimate concerns over the use of minimization in practice, or the fact that these concerns generally center on a lack of allocation concealment [8,9]. How can allocation concealment be called into question when we have just established that upcoming allocations cannot be predicted? Is this not the very essence of allocation concealment?

The author has previously [6] noted that investigators have allocation discretion. That is, screening a patient for a trial does not obligate the investigator to randomize that patient, either immediately or at all. Rather, investigators can defer enrollment, deny enrollment, or derail enrollment by projecting doubts that cause the patient to decline entry into a trial for which he or she is qualified based on the entry criteria. This means that investigators still have discretion even after they have identified a specific candidate patient who may or may not be enrolled. Let us imagine that every other aspect of allocation concealment is perfect, so that just before a candidate patient is identified, the investigator has no idea which treatment is due up next. This means that the investigator has no reason to prefer an older or younger, a healthier or a sicker patient.

However, once a candidate patient is identified, everything changes. The key variables that are to be balanced by the minimization algorithm will be observed, and with this, the investigator can now determine (or at least predict) the next allocation. Whether or not there is any advantage to be gained in doing so, and whether or not any investigators actually would do so, these are immaterial for our purposes. It is enough for us to note that the next allocation can be predicted early enough that the investigator can still put a halt to it. Hence, there is no true allocation concealment, because the source of uncertainty regarding the next allocation stems from uncertainty of the next set of patient characteristics to walk through the door. As noted, once the patient enters, and these baseline characteristics are observed, the next treatment assignment can be discerned.

Conditional & unconditional allocation concealment

The preceding discussion makes clear that minimization has a valid claim to allocation concealment in one sense, while clearly precluding the possibility of allocation concealment in another very real sense. So we are confronted with the reality that no single concept of allocation concealment will suffice, and this is true more broadly for any adaptive randomization technique. The logical conclusion, then, is that we need not one but two concepts of allocation concealment, so as to handle adaptive randomization techniques. We will refer to the standard concept of allocation concealment as unconditional allocation concealment. That is, prior to identifying a candidate patient the investigator cannot predict the upcoming allocation. Our new concept is conditional allocation concealment. That is, even after identifying a candidate patient, the investigator still cannot predict the upcoming allocation. With these definitions, it is clear that minimization is consistent with unconditional allocation concealment, but is not consistent with conditional allocation concealment.

An illustrative but hypothetical example

To keep the example simple, imagine first that gender is the only variable to be balanced by the use of minimization in this two-arm parallel trial comparing A to B. Suppose, further, that when there is a gender imbalance, the allocation is deterministic. That is, if there are more females already allocated to A, and the next patient is a female, then she will be allocated to B with probability one. If there are more males already allocated to A, and the next patient is a male, then he will be allocated to B with probability one. If there is balance for the gender of the next patient, then this patient will be allocated by the equivalent of a coin toss. Clearly, there are other algorithms that could be used when using minimization,

and just as clearly, there are other adaptive procedures besides minimization, but this will suffice to make the point.

Suppose that for whatever reason the investigator would like to subvert the randomization and get patients with more severe disease to receive A, and patients with less severe disease to receive B. Disease severity is certainly observable (in this example), whether it ends up as a measured variable or not. Note that while there is ample evidence that this type of selection bias does in fact occur in practice (e.g., [10] and Chapter 3 of [6]), our concern would be justified even if we had no evidence that this ever occurs in practice. Loopholes must be closed even if we are not aware of their being exploited, because the fact that they could be exploited is enough to call the validity of the research into question. So our concern here is with the possibility of cheating, rather than with the frequency with which cheating occurs, or how many investigators would engage in this type of cheating if presented the opportunity to do so.

Note also that the Markovian nature of the minimization algorithm lends itself to cheating of this type. That is, the allocation probabilities for the next patient, conditional on the entire enrollment history up to this point, depends on only the present configuration, so it is easy to keep running totals, and this facilitates cheating. One could, however, modify the algorithm so that it would no longer hold this property. For example, the algorithm might also track trends, and weigh more recent allocations more heavily than earlier ones. That is, one would seek to ensure not only overall balance but also 'local' balance, perhaps over the past ten allocations. The whole point of the algorithm is to ensure that there are no trends, so this is a moot point in our example with only one balancing covariate and deterministic allocations to force balance. But with less extreme biasing probabilities and/or multiple balancing covariates, it would in fact be possible to see trends in any one balancing covariate, so these might be taken into consideration by a more complicated, and less predictable, minimization algorithm.

Be that as it may, for our purposes, the only relevant variables are patient gender (male or female), disease severity (mild or severe), treatment allocation (A or B) if randomized, and an indicator (yes or no) of whether or not the patient actually gets randomized. Suppose that the first patient is a female with mild disease, and gets randomized, and allocated to A. Then the next patient is a male with severe disease, and also gets randomized, and allocated to A. Given the algorithm, we now know that the next patient to be randomized, whether male or female, will get allocated to B. Recall that the investigator would like to subvert the randomization and get patients with more severe disease to receive A, and patients with less severe disease to receive B. This will drive the decisions to enroll certain patients or not, and this phenomenon persists even when the minimization allocations are not deterministic.

A recommendation on how to randomize

It is widely accepted, though generally unspoken, that randomization is randomization, and it matters little how it is actually done or, for that matter, even if it is done at all, as long as the claim is made. Berger and Bears [11] noted a rather startling discrepancy between claims of randomization and verifiable randomization in trials that are considered (rightly or wrongly) to have been randomized. Quite often there is no information at all regarding how the randomization was conducted, and while this should raise a red flag for journal editors and reviewers, in practice it rarely does.

Some studies have been known to use alternation between the two treatments, and yet are still labeled as randomized. However this is most certainly not randomized, nor is minimization in its pure form. Both procedures preclude the possibility of allocation concealment, since in each case it is clear what the next subject will receive (in one case, this does not depend on the patient characteristics, and in the other it does, but given those observable patient characteristics, one can always deduce what treatment will be assigned). So whereas alternation balances the numbers of patients allocated to each treatment group, it was not result in good covariate balance.

At the other end of the spectrum is complete randomization, in which there is no relationship between one allocation and any other; the allocations are completely independent, as if a coin were being tossed for each allocation (although in practice this is not how the procedure is implemented). As noted by Berger, Ivanova, and Deloria-Knoll [5] and Berger [6], this is the ideal procedure from the perspective of preventing the prediction of future allocations based on knowledge of past ones. However, this complete randomization may result in treatment groups of vastly different sizes. For this reason, it is common, and appropriate, for some type of restricted randomization to be used. Alas, there are many different types of restrictions that can be used, and that are used in practice, and some are more suitable than others. The key attributes upon which a randomization scheme will be measured, or should be measured, are predictability and balance. As noted, alternation is ideal for balance, in that it ensures equally sized groups (as long as it is not terminated prematurely), but less suitable for prediction, whereas complete randomization is entirely unpredictable, but often leads to imbalance.

In fact it is true in general that there is a trade-off between balance and predictability, in that forcing better balance inevitably leads to greater prediction (or, phrased differently, compromised allocation concealment). This

may be seen most readily by considering varying the block size for use with permuted block randomization. Larger block sizes force fewer returns to perfect balance, and so are less predictable, but also allow for more imbalanced treatment groups (in terms of group size). Conversely, smaller block sizes force more returns to perfect balance (this return to perfect balance is forced at the end of each block), and so are more predictable, but also do better at creating treatment groups that are balanced for size.

The maximal procedure [5,6] is a better procedure than permuted blocks, in the sense that if one selects the maximally tolerated imbalance to match that of the permuted block (in other words, half the block size), then the maximal procedure is less predictable. Moreover, unlike with permuted blocks, with the maximal procedure it is also generally not possible to predict when prediction will be possible. Note that varying the block size does not render permuted blocks any more appropriate [6].

With blocks known to be of size four, for example, it is known in advance that the fourth, eighth and twelfth and every fourth allocation will be predictable. The maximal procedure will allow some predictable allocations, but which ones cannot be determined in advance, and instead must be revealed as the allocations unfold. Along similar lines, the benefit of the maximal procedure over permuted blocks increases even more when unmasking is only partial. Consider, for example, a situation in which the first five allocations are all successfully masked (meaning that the identity of the treatments allocated remain unknown until the end of the trial), but then the sixth and seventh are both unmasked and seen to be the active treatment. Then, without even knowing the earlier allocations, one can, if using permuted blocks of size four, still determine that the eighth allocation to be control. With the maximal procedure, this is not possible. If at any point during the randomization partial masking is enough to make a 'card counter' lose track, then it will be almost impossible for that card counter to get back on track and make useful predictions.

The big stick procedure [7] is quite similar to the maximal procedure, and differs only with respect to the probabilities of allocation to either treatment before the reflecting barrier is reached. That is, the maximal procedure can be thought of in the following way (and can even be conducted in this way, if no more efficient algorithm is found). Enumerate all possible allocation sequences of the desired length (i.e., the specified sample size of the trial). For each of these potential allocation sequences, compute the largest numerical difference ever attained between the number of patients allocated to one treatment group and the number of patients allocated to the other treatment group. Delete all sequences that exceed the specified maximally tolerated imbalance (MTI). In general, if one would contemplate using permuted blocks of a fixed block size, then the MTI for use with the maximal procedure will be half this block size. Randomly select one sequence from among those that remain.

The maximal procedure will force a return towards balance when the imbalance reaches the MTI. If, for example, the MTI is two, and the sequence begins AABAB, then after the second and fourth allocations the imbalance will be two, thereby reaching the MTI, and B will be forced as the third and fifth allocations. But, unlike with permuted blocks of size four, the fourth allocation will not be forced, and is allowed to be A again. It is more likely that B will be the next allocation, but the conditional probability of B as the fourth allocation, given the initial sequence AAB, is still less than 1 (albeit more than 0.5). This means that the fourth allocation can be predicted, but not with certainty. The big stick procedure uses the same set of potential allocation sequences as the maximal procedure does, but all allocation probabilities are 0, 1, or 0.5. That is, the big stick procedure does the same thing at the reflecting boundaries when the MTI is reached, but it differs in the other allocations, in that these other allocations not only do not force returns to balance, but in fact, unlike the maximal procedure, they do not even encourage returns to balance. It is akin to tossing a coin, as long as the MTI is not reached so either A or B can be allocated next.

We see, then, that the key difference between the maximal procedure and the big stick procedure is in how they handle allocations that are not forced. By encouraging returns to balance, the maximal procedure will tend to reach the MTI less often, and hence will be expected to produce fewer deterministic (predictable) allocations. So by this metric, the maximal procedure is to be preferred. However, the big stick procedure will have fewer allocations that allow one to make some sort of prediction, even if not perfect. So by this metric, the big stick procedure is to be preferred.

It is not unusual, when comparing two procedures, to find that one is better by one metric and another is better by another metric, and this may leave one unclear as to which one to use in which situations. In our case, we see readily that the crux of the matter is prediction with less than certain odds of being right. If we somehow knew that investigators would not try to strategically select patients for one treatment group or the other unless they were certain of the next allocation, then we would want to use the maximal procedure. If, however, we had reason to believe that the investigators would do so even with uncertainty, as long as the probabilities favored them, then we would want to use the big stick procedure. Obviously this is not a question that one could ask an investigator, and indeed some bias will be unintentional.

One saving grace to alleviate this dilemma is the distinction between masked and unmasked studies. As noted by Berger [6], we can trust that unmasked studies are unmasked to a much greater extent than we can trust that masked studies are masked. There are so many ways for unmasking to occur, and so many ways for this to escape detection, that one can never really say for sure that a future study will be successfully masked, or that a past one was successfully masked. With this is mind, we recognize masked studies as studies that are planned as masked, as opposed to studies that are necessarily successfully masked. The distinction, then, is between imperfectly masked trials and unmasked trials.

In unmasked trials, the entire allocation sequence is observed as it unfolds, so that at any point in time, one may know all prior allocations. With knowledge of the randomization procedure, therefore, one can also compute the probabilities of the next allocation, and will be in a position to select, for example, a healthier patient if the active treatment is more likely to come up next, or a sicker patient if the control is more likely to come up next. The big stick procedure might be ideal in unmasked trials to prevent this type of selection bias that would operate with less than certainty. In ostensibly masked trials, however, it might be more difficult for an investigator to engage in this type of selection bias, since, we would hope, enough of the past allocations would be unknown to render them uncertain as to which treatment is more likely to come up next, except for when a long string of allocations to the same treatment occurs.

For example, if the MTI is two, and at a certain point during the randomization process the investigator has no idea how many patients have been allocated to each treatment (other than that the two numbers must be within two of each other), and then observes four allocations in a row all to the control group, then the investigator can deduce that prior to this run there has been two more patients allocated to the active group, but now, of course, there are two more allocated to the control group. The MTI has been reached and, unlike four allocations ago, now is known to have been reached. The reflecting boundary kicks in, and the next allocation must be to the active treatment, so the investigator would be in a position to recruit a healthier patient.

This type of selection bias, operating with certainty, is our primary concern in (ostensibly) masked trials, and since the maximal procedure will tend to produce fewer deterministic allocations than the big stick procedure will, we would recommend that the maximal procedure be used in trials planned as masked. In summary, we recommend the maximal procedure for a masked and unstratified study in a single center, the big stick procedure an unmasked and unstratified study in a single center, and a complex combination (varying the MTI too) of the two for stratified and/or multicenter trials (which of course would stratify by center).

Berger, Grant, and Vazquez [12] applied Rosenbaum's earlier work [13] on sensitivity to the problem of randomizing and concluded that it is ideal to vary the approach across strata within a study, but left open the possibility of using permuted blocks and minimization in some strata. We would prefer to stick exclusively with the big stick procedure and the maximal procedure, but vary which is used, and the MTI. One may well ask at this point how one would randomize the centers and/or strata (which may be defined based on factors other than centers, including age, gender, and/or disease severity) to determine which are conducted with the big stick and which with the maximal procedure, and what MTI is to be used for each.

We are not too particular on how to determine which strata use the big stick and which use the maximal procedure. It would probably suffice to use unrestricted randomization (toss a coin for each one), but the keys are that:

First, this coin be biased towards the maximal procedure for masked studies and towards the big stick procedure for unmasked studies and;

Second, the results of this initial randomization (of strata and/or centers) be kept confidential until after the study.

The biased coin may follow Efron [15] and use one specified probability for masked trials and another for unmasked studies, or it may instead follow Wei [16] and allow the biasing probability to depend on an assessment of how unmasked the study is thought (at the outset) likely to be.

As for the MTI, we might like to see it varied not only across strata, but also within strata, but always in a decreasing fashion. That is, we envision eight allocation approaches, to be used among the strata within a study. The parameters are the varying MTIs to be used. So, for example, a stratum using the maximal procedure with (4, 3, 2) will initially have an MTI of four, and after roughly 33% of the patients have been allocated this will switch to an MTI of three, and when roughly 67% of the patients have been allocated this will switch again to an MTI of two, so that the final group sizes can differ by no more than two. These eight approaches are:

1. Maximal (4, 3, 2);
2. Maximal (4, 3);
3. Maximal (4, 2);
4. Maximal (3, 2);
5. Big stick (4, 3, 2);
6. Big stick (4, 3);
7. Big stick (4, 2);
8. Big stick (3, 2).

In practice, of course, the switches are built in to the allocation sequences, and not left until the appropriate proportions of patients have been randomized. The precise switch point is also varied, so that the first may be anywhere between 25 and 35% of accrual completed, and the second may be anywhere between 65 and 75% of accrual completed. This complex approach would be nearly impossible to predict and would also result in comparable group sizes all throughout the study, and especially at the end.

Conclusion

In chess terminology, an opening or a style of play is said to have sharp lines if the path to victory is both narrow and treacherous, with mine fields everywhere. The same might be said for the conduct of a valid clinical trial. There are many ways to get it wrong, and, unfortunately, very few ways to get it right. This applies even to an aspect as fundamental as the randomization itself. Flawed randomization methods are the norm, and it is rare to see any serious thought put into the methods that are used. Often it comes down to which method the researchers prefer, and then there is no scrutiny.

Flawed randomization methods, however, have real consequences, in that they subvert the randomization so that the treatment groups may be systematically different, and this, of course, precludes the possibility of unbiased research. We have proposed a novel variation of allocation concealment to better place various randomization methods in context. This is but one of several tools that can be used to assess the validity of a randomization method, both in general and as it pertains to any specific trial. We can only hope that researchers will improve future trials, both in general and regarding the ways they choose to randomize. It would appear that there is no place for permuted blocks given the ready availability of uniformly better methods.

At the very least, a compelling argument would be needed to justify permuted blocks, and appeal to convention or precedent is not a compelling argument. Nor is the convenience of the research team. Two common arguments bear discussion. One is that central randomization takes care of the problem, because (in a multicenter trial) investigators at one center have no knowledge of what is happening at any other center. The other is that permuted blocks of size four may be used safely as long as their use is not revealed in the protocol, because the investigators would need to be aware of this to act on it. Each argument has its appeal; however, each is incorrect. In multicenter studies, even when central randomization is used, this randomization is still almost always stratified by center. Hence, each center has its own allocation sequence, which can be predicted without knowledge of the unfolding of any other allocation sequence, and without actually being directly observed. And it is clearly a mistake to equate the withholding of the randomization procedure from the protocol with the inability of investigators to determine this procedure, either from past experience with the same sponsor or from noticing that after every four patients there always seems to be perfect balance.

There is evidence that prediction of this sort happens. For example, Chapter 3 of [6] lists no fewer than 30 trials with strong indications that this has occurred, and Fayers and King [14] described another. We might regard these 31 trials as the tip of the iceberg when we consider trials with a selection bias that have not been identified. One cannot expect to find trials fitting the bill just by searching 'selection bias' in Pub Med or directly asking investigators.

It seems safe to say that selection bias is a real problem, not a hypothetical one, and as outlined above, the assurances offered by central randomization and sanitizing the protocol of the randomization method are rather misguided. Hence, we still need better randomization methods to be used in practice, even when the randomization is central.

Future perspective

There has been much cutting edge work recently in the proposal of novel methods of randomization designed to prevent attempts at prediction, and, ultimately, ensure balance. After all, it is crucial that the comparison groups be as comparable as possible. One would expect better penetration of these novel methods into actual trials over the next decade or so, as more attention is paid to this key issue in trial quality. To the extent that this does come to pass, we can expect more rigorous trials overall (i.e., one would hope that greater attention to this issue might translate into greater attention also to other key issues of trial quality), and this will mean that trials will produce results that better reflect reality, so that there will be a better chance of distinguishing promising treatments from those that are not sufficiently efficacious to justify their use.

Financial & competing interests disclosure

The author has no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

- Trials represent our best chance to rigorously evaluate treatments.
- To be valid, trials need to ensure that the comparison groups are actually comparable to each other at baseline.
- One threat to baseline balance is flawed randomization which allows for prediction of future allocations.
- There are existing randomization methods that control selection bias, but these tend not to be used in practice.
- One barrier to the use of the better randomization procedures may be insufficient clarity that they are in fact superior to the methods currently in use.
- Current evaluation methods tend to treat standard and adaptive randomization procedures as separate and distinct.
- We have proposed a new method of evaluation that places the two types of randomization on an equal playing field, so that they can be compared directly.

References

Papers of special note have been highlighted as:

- **of interest**

1 Cohen MM, Kaufert PA, MacWilliam L, Tate RB. Using an alternative data source to examine randomization in the Canadian national breast screening study. *J. Clin. Epidemiol.* 49, 1039–1044 (1996).

2 Berger VW, Christophi CA. Randomization Technique, Allocation Concealment, Masking, and Susceptibility of Trials to Selection Bias. *JMASM* 2(1), 80–86 (2003).

3 Berger VW. Is Allocation Concealment a Binary Phenomenon? *Med. J. Australia* 183(3), 165 (2005).

4 Berger VW, Do AC. Allocation Concealment Continues To Be Misunderstood. *J. Clin. Epidemiol.* 63(4), 468–470 (2010).

5 Berger VW, Ivanova A, Deloria-Knoll M. Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. *Stat. Med.*22(19), 3017–3028 (2003).

- **Formulates the proper framework for evaluating randomization methods, balancing control of prediction (selection bias) against control of chronological bias, and also proposes the maximal procedure, which remains one of the best randomization methods available.**

6 Berger VW. *Selection Bias and Covariate Imbalances in Randomized Clinical Trials.* John Wiley & Sons, Chichester, UK (2005).

- **Offers a fairly complete overview of issues pertaining to randomization.**

7 Soares JF, Wu CFJ. Some restricted randomization rules in sequential designs. *Commun. Stat. Simul. Comput.* 12, 2017–2034 (1983).

8 Berger VW. Minimization, by its nature, precludes allocation concealment and invites selection bias. *Contemp. Clin. Trials* 31, 406 (2010).

9 Berger VW. Minimization: not all it's cracked up to be. *Clin. Trials* 8, 443 (2011).

10 Berger VW, Weinstein S. Ensuring the comparability of comparison groups: is randomization enough? *Control. Clin. Trials* 25(5), 515–524 (2004).

11 Berger VW, Bears J. When can a clinical trial be called 'randomized'? *Vaccine* 21, 468–472 (2003).

12 Berger VW, Grant WC, Vazquez LF. Sensitivity designs for preventing bias replication in randomized clinical trials. *Stat. Methods Med. Res.* 19(4), 415–424 (2010).

13 Rosenbaum PR. Design Sensitivity in Observational Studies. *Biometrika* 91, 153–164 (2004).

- **Clarifies that the best evidence comes from studies that are not replicates of each other but rather vary the challenges to the hypotheses.**

14 Fayers PM, King M. A highly significant difference in baseline characteristics: the play of chance or evidence of a more selective game? *Qual. Life Res.* 17, 1121–1123 (2008).

- **Illustrates a trial with flawed randomization.**

15 Efron B. Forcing a sequential experiment to be balanced. *Biometrika* 58, 403–417 (1971).

16 Wei LJ. The adaptive biased coin design for sequential experiments. *Ann. Stat.* 6(1), 92–100 (1978).

- **Builds on Efron's earlier work to expand biased coin randomization so as to allow for a biasing probability that depends on the degree of imbalance.**