

Randomization in clinical trials: can we eliminate bias?

Clin. Invest. (2013) 3(1), 37–47

Randomization plays a fundamental role in clinical trials. While many modern clinical trials employ restricted, stratified or covariate-adaptive randomization designs that pursue balance in treatment assignments and balance across important covariates, some clinical trials call for response-adaptive or covariate-adjusted response-adaptive (CARA) randomization designs to address multiple experimental objectives primarily related to statistical efficiency and ethical considerations. In this paper, we elicit key principles of the well-conducted randomized clinical trial and explore the role of randomization and other important design tools in achieving valid and credible results. We give special attention to response-adaptive and CARA randomization designs, which have a firm theoretical basis, but are more complex and more vulnerable to operational biases than traditional randomization designs. We conclude that modern advances in information technology, rigorous planning, and adherence to the key principles of the well-conducted clinical trial should enable successful implementation of response-adaptive and CARA randomization designs in the near future.

Keywords: accidental bias • allocation concealment • covariate-adaptive
• covariate-adjusted response-adaptive • covariate imbalances • masking
• selection bias • stratification

Randomized clinical trials

The randomized clinical trial (RCT) is recognized as the most credible research design for clinical investigation [1,2]. The goal of the RCT is to achieve valid comparison of the effects of an investigational treatment or treatments with the control treatment (standard of care) in the target patient population. The hope is to demonstrate that the investigational product is safe and efficacious in treating the disease and can lead to an improved quality of life. In a RCT, eligible patients enroll sequentially and must be allocated immediately to one of the treatment groups. In order to achieve valid and precise treatment comparison, the treatment groups must be 'comparable' in terms of important patient characteristics. How should the treatment assignments be made to ensure such comparability? Randomization forms the basis of a clinical trial design and is used to ensure that statistical inference in the end of the trial is valid [3]. Books have been written on theoretical aspects of randomization and its applications in clinical trials [4,5]. However, is randomization by itself sufficient to achieve the goal of a RCT? In this paper, we shall explore the role of randomization and other key features of the well-conducted RCT in combating different experimental biases and contributing to credibility of the trial conclusions.

Bias & its control in randomized clinical trials

Bias in the RCT can arise from multiple sources. The International Conference on Harmonization E9 (ICH E9) guidance describes bias in a clinical trial as the

Oleksandr Sverdlov¹
& **William F Rosenberger*²**

¹Translational Sciences, Novartis Pharmaceuticals Corporation, NJ, USA

²Department of Statistics, George Mason University, 4400 University Drive MS 4A7, Fairfax, VA 22030, USA

*Author for correspondence:

Tel.: +1 703 993 3645

E-mail: wrosenbe@gmu.edu

**FUTURE
SCIENCE** part of **fsg**

“systematic tendency of any factors associated with the design, conduct, analysis and evaluation of the results of a clinical trial to make the estimate of a treatment effect deviate from its true value” [101]. The ‘systematic tendency’ implies that if a trial were to be repeated (even hypothetically) the bias would still be present and would still invalidate the results. The ICH E9 guidance distinguishes operational and statistical bias; the former refers to bias introduced during the conduct of the trial, whereas the latter refers to bias introduced in the design, analysis and evaluation of the results. It is also important to distinguish bias from the error caused by sampling [6] that arises from studying the finite sample of patients to estimate the true treatment effect.

In the literature, several authors provide classifications of various types of biases in clinical studies [6–10]. Here we first briefly describe major common biases that can arise in a RCT during the design, conduct, analysis, interpretation and reporting of the results, and after that we will focus on specific biases that randomization is intended to mitigate.

- **Design bias:** this can arise at the planning stage of a clinical trial. It may be due to a wrong choice of the trial design (a design that does not answer the research questions), inadequate sample size, incorrect assumptions of the treatment effect and/or variability, incorrect choice of patient population and inclusion/exclusion criteria, wrong assumptions about enrollment and dropout patterns, and so on. Note that the issues described herein are not examples of systematic biases as described in the ICH E9 guidance [101].
- **Conduct bias:** this can arise during the conduct of a clinical trial. The major subtypes of conduct bias include: ascertainment bias due to knowledge of which treatment each study participant is receiving; selection bias due to the investigator’s attempt to guess the treatment assignment and selectively enroll study participants; and accidental bias due to impact of important unknown covariates on the primary outcome. It is important to note that biases described in this category are hard to quantify and adjust for in the statistical analysis.
- **Analysis bias:** this can occur at the data analysis stage due to incorrectly chosen statistical methodology, model misspecifications, improper handling of missing data, poor quality of analysis datasets, and so on.
- **Interpretation and reporting bias:** this can occur when the trial results are incorrectly interpreted and disseminated to the broader community. Frequently encountered are: significance bias (which refers to confusion of statistical significance with clinical significance), bias related to competing interests

(financial sponsorship of the trial), and publication bias (selective publication of the trials with positive results; see [11]).

The aforementioned list of biases is by no means exhaustive. Substantial effort is needed to plan and implement the trial so that reliable conclusions can be reached. Randomization is necessary, but not sufficient in mitigating all possible biases in the study. However, the carefully implemented randomization design can mitigate or minimize certain biases that otherwise can have major detrimental impact on the validity and integrity of the trial results.

Which operational biases can be mitigated by the proper use of randomization?

In summary, the merits of randomization are that, in conjunction with other important design techniques, such as masking and allocation concealment, it can mitigate major operational biases that negatively impact the study results [2,101]. The following will briefly describe randomization, masking, and allocation concealment in the context of a RCT.

Randomization refers to generation of a sequence of treatment assignments by means of some known random mechanism (e.g., by a flip of a coin). Various randomization procedures with established statistical properties are available for use in practice [4,5]. One should distinguish advance randomization and adaptive randomization procedures. With advance randomization, a treatment allocation sequence can be pre-generated before any subject is enrolled into the study by enumerating a set of all possible allocation sequences and selecting one randomly with known probabilities [12]. Examples of advance randomization are the popular permuted block design, the maximal procedure [13] and the Hadamard randomization [14], to name a few. Rosenberger and Lachin (in Chapter 3 of [4]) describe restricted randomization procedures, including Efron’s biased-coin design [15] and its generalizations, many of which can be cast as advance randomization. In contrast, adaptive randomization does not use pre-generated treatment allocation sequences; instead, treatment randomization probabilities are modified through the course of the trial based on accumulated information on treatment assignments, patient responses, and/or covariates to achieve specific experimental objectives while maintaining the validity and integrity of the trial results [16]. Within the class of adaptive randomization, one should distinguish covariate-adaptive, response-adaptive and covariate-adjusted response-adaptive (CARA) randomization. Adaptive randomization provides greater flexibility and frequently it can achieve trial objectives more efficiently than advance randomization, but it is also associated

with additional logistical complexities and sometimes it can be more vulnerable to experimental bias.

Masking in a clinical trial refers to a process that attempts to keep the treatment assignments unknown or easily ascertained by those who are ‘masked’ [17,18]. Allocation concealment is the prevention of knowledge of a given treatment allocation until after it is executed, to avoid selective enrollment of patients into the trial [19].

Let us discuss in detail how randomization, masking and allocation concealment can help mitigate experimental biases during the conduct of the trial. We focus on four features that may potentially contribute to bias during the trial conduct: impact of the knowledge of which treatment each participant is receiving (ascertainment bias); impact of an investigator on the process of treatment assignments (selection bias); impact of important known covariates of a patient on response to the treatment; and impact of important unknown covariates on the primary outcome (accidental bias).

■ Ascertainment bias

After an eligible patient is randomized into the trial, ideally, the patient, the investigator, the healthcare provider and the outcome assessors should not be affected by the knowledge of the treatment assigned. Ascertainment bias refers to systematic distortions due to the knowledge of which intervention the patient is receiving. For instance, a patient may be more likely to drop from the study or not adhere to the treatment regimen if he knows that he is assigned to placebo; an investigator may provide different concomitant medication care to patients on active drug and placebo; outcome assessors may be biased in the assessment of patient outcomes based on their preconceived expectations of treatment effects. The most efficient technique to mitigate ascertainment bias in a RCT is masking. One should distinguish open-label trials in which “*the identity of treatment is known to all*”; single-masked trials in which, “*the investigator and/or his staff are aware of the treatment but the subject is not, or vice versa*”; and double-masked trials in which, “*neither the study subjects nor any of the investigator or sponsor staff who are involved in the treatment or clinical evaluation of the subjects are aware of the treatments received*” [101]. The double-masked trial is regarded as the optimal approach. As noted by Friedman *et al.*, “*A clinical trial should, ideally, have a double-blind design to avoid potential problem of bias during data collection and assessment. In studies where such a design is impossible, a single-blind approach and other measures to reduce potential bias are favored*” [3]. Some useful approaches to maintain double-masking throughout the trial include the use of placebo controls, the use of ‘double-dummy’ technique when the

two treatments cannot be made identical and providing training/education to all personnel involved in the clinical trial on the concept of masking and its importance for the integrity of the trial results [3,6,20].

■ Selection bias

Some authors hold the view that selection bias is the major type of bias in clinical trials [12,21,22]. Selection bias may not be an issue in a randomized double-masked trial where treatment allocations are carefully concealed throughout the trial. Allocation concealment supplements randomization and masking in that it keeps study investigators unaware of any upcoming treatment allocation. Some strategies for allocation concealment include use of sealed envelopes [19], use of central interactive voice response systems and centralized computer- or web-based allocation systems [23].

In practice, however, successful masking and successful allocation concealment are not always possible. In an unmasked study an investigator may be able to guess the treatment assignments of future patients based on knowing the treatments assigned to the past patients and selectively enroll a patient who, in their opinion, will benefit most from the given treatment. Selective patient enrollment can negatively impact the study in a number of ways. One consequence of selection bias is explained by Longford: if treatment effects are heterogeneous in the population, selection bias can lead to biased estimation of the mean treatment effect, under-represented population treatment heterogeneity, or both [24]. Another negative consequence is systematic covariate imbalances. For instance, one of the treatment groups may have a greater number of ‘sicker’ patients and therefore the observed treatment effect will be biased. Systematic covariate imbalances can inflate type I error [25,26], and they can undermine the integrity of the trial [21]. The mechanism of selection bias that results from prediction of an upcoming treatment assignment is referred to by Berger as ‘third-order selection bias’ [21]. To minimize predictability of treatment assignments it is important to avoid using very restrictive randomization procedures, such as the permuted-block designs with small block sizes. The maximal procedure can be used to generate the least restrictive randomization procedure subject to a constraint on the maximum tolerated imbalance [13]. Efron’s biased-coin design [15] is another valuable fully randomized procedure with established exact statistical properties [27]. Statistical techniques are available to test for selection bias in a RCT [26,28] and adjust for observable selection bias [29]. Rosenberger and Lachin (in Chapter 6 of [4]) quantify susceptibility of different randomization procedures to selection bias using Blackwell and Hodges’s model for selection bias [30] and make recommendations for practice.

We have described selection bias as the selective enrollment of patients into the study. However, selection bias can be also introduced at the data analysis stage – by postrandomization exclusions, either through the loss or withdrawal of randomized patients or through the use of the non-intention-to-treat analysis populations. Such exclusions represent, perhaps, the largest potential for bias, even in properly randomized trials with adequate allocation concealment and masking; for example, differential loss can occur even in otherwise well-conducted trials if, say, the active drug is not well-tolerated by certain patients whereas patients randomized to placebo might drop out for lack of perceived benefit. It should be emphasized that randomization does nothing to prevent such a bias and only sufficient attention and resources committed to active retention will do.

■ Impact of known covariates

It is recognized that randomization can promote balance of important known and unknown patient baseline characteristics (covariates) across treatment groups, which is a major prerequisite for valid treatment comparison [101]. Recent literature reports show that non-randomized trials and observational studies tend to produce biased estimates of intervention effects compared with randomized trials, likely due to confounding of the effects of treatments and prognostic factors [31,32]. By contrast, in a well-conducted RCT, there is high confidence that any observed treatment difference should be attributed to the treatment effects, not to the effects of the study patients [33]. Note, however, that any observed difference might also be attributable to chance, which can never be completely ruled out [34]. Such chance differences are really the results of random imbalance in some (known or unknown) patient characteristics. How can one be assured that randomization has satisfactorily achieved its purpose and produced comparable treatment groups? Since there are different ways to measure similarity of treatment groups in terms of covariate profiles, what does ‘comparability’ exactly mean? Senn makes an important point that random covariate imbalances do not compromise the validity of inference [35]. An unbiased treatment comparison is achieved by proper accounting of covariate information through covariate-adjusted analysis using regression modeling.

A controversy about the role of covariates in the design of clinical trials is elicited by Rosenberger and Sverdlov [36]. They raise the question of why attempts should be made to prospectively balance the distributions of important covariates across treatment groups? For trials in which primary outcomes follow a homoscedastic linear regression model, a design that achieves balanced treatment assignments, both overall in the trial

and across covariate profiles will maximize power and efficiency of treatment comparison. Therefore, randomization procedures that force balance across covariates are justified from the standpoint of statistical efficiency in such trials. While complete randomization and restricted randomization procedures balance covariate distributions for large samples, in trials with small or moderate sample sizes covariate imbalances may be substantial. Stratified randomization can ensure treatment balance by using permuted block randomization within mutually exclusive strata formed by the levels of selected covariates, which can improve efficiency of small trials and can facilitate subgroup and interim analyses for large trials [37]. Another advantage of stratified randomization is that some strata can be dropped from the analysis without affecting the integrity of randomization in other strata [4].

For a small number of strata, stratified randomization works well. However, for a larger number of strata, some strata may be empty or may contain very few patients. If some strata are empty, then there is nothing to balance for; however, if a stratum contains a few patients, then within a given block in the stratum only initial treatment allocations will be utilized and the overall treatment imbalance may be substantial [38]. To address this issue, covariate-adaptive randomization procedures (sometimes called minimization procedures), have been proposed. With a covariate-adaptive randomization procedure, an eligible patient is randomized to a treatment with probability conditional on the full history of previous patients’ treatment assignments and covariates, and the covariates of the current patient [4]. The goal is to sequentially balance the distributions of covariate profiles across treatment groups while maintaining randomization. Rosenberger and Sverdlov [36] give an overview of covariate-adaptive randomization methods and distinguish stratified randomization, marginal covariate-adaptive randomization [39,40], and optimal design-based covariate-adaptive randomization [41]. They note that covariate-adaptive randomization leads to nearly optimal allocation when responses follow a linear regression model with constant variance and no treatment–covariate interactions. They further point out that marginal covariate-adaptive randomization procedures may be advantageous over stratified randomization when the goal is to achieve balance within a very large number of covariate margins. Recent literature reviews show increased popularity and application of covariate-adaptive randomization methods in clinical trials [42,43]. It is generally acknowledged that any stratification or minimization variables used in the design should be also adjusted for in the analysis for the inference to be valid [35,44]. In fact, proper accounting for covariates in the analysis

“may make a more valuable and instructive contribution to inferences about treatment effects than only balancing them” [45]. Operationally, covariate-adaptive randomization should be performed by a centralized allocation unit using interactive voice response systems [23,46,47] that must be validated to avoid errors in the allocation, and every effort should be made to retain the double-masked status of the trial. A random element must be present in a covariate-adaptive randomization procedure to mitigate possibility of selection bias [48–50].

Recent research work has been substantial both on exploration of theoretical properties of existing covariate-adaptive randomization designs [51,52] and the development of new methods [53–56]. Undoubtedly, the current body of knowledge of covariate-adaptive randomization is far from being mature and there is quite a lot more to investigate.

Overall, one should acknowledge that covariate-adaptive randomization can improve efficiency and interpretability of the results in small and moderate sample clinical trials employing homoscedastic linear regression models where balance over a very large number of important baseline covariates is warranted. However, in clinical trials with heteroscedastic and nonlinear models for the primary outcomes, balanced allocation is not necessarily optimal and alternative randomization procedures (such as CARA randomization) may be preferred [36].

■ Accidental bias

The term ‘accidental bias’ was introduced by Efron, and it refers to bias due to factors unforeseen at the design stage [15]. Suppose there is a covariate that an investigator is unaware of, that has a strong impact on patients’ outcomes. If this covariate is omitted from the statistical model in the analysis, the estimated treatment effects will be confounded with the covariate effect and the treatment comparison will be biased. Randomization tends to mitigate bias from unknown covariates. Efron’s model for accidental bias quantifies the magnitude of severe covariate imbalances for various randomization procedures. Different randomization procedures can be ranked in terms of their susceptibility to accidental bias [4]. In essence, the more random a procedure is, the less susceptible it is to accidental bias. Complete randomization for which every treatment allocation in a two-arm trial is made with probability 0.5 is most random and it is least susceptible to accidental bias. In contrast, a randomization procedure that generates a sequence of treatment assignments with some particular periodicity increases the likelihood of periodicity in the sequence of covariate values and may potentially lead to covariate imbalances. For example, consider a truncated binomial design [30] that for a sample size n makes allocations

with probability 1/2 until one of the groups achieves the target of $n/2$ patients, and then assigns the remaining patients to the other treatment with probability 1. Such a design is highly susceptible to accidental bias when n is large due to a high likelihood of a sequence of deterministic assignments in the tail of the randomization sequence [57]. To avoid this feature, one can apply restricted randomization procedures in blocks to ensure that balance in treatment numbers is achieved throughout the course of the trial. Many restricted randomization procedures described in [4] handle accidental bias fairly well, both in theory and as shown through simulations. Rosenberger and Lachin (in Chapter 5 of [4]) note that Efron’s model of accidental bias quantifies the magnitude of severe covariate imbalances, not its risk, and they conclude that, in practice, accidental bias should be less of a concern than selection bias [4].

Which biases should be controlled at the data analysis step?

Data analysis is the crucial component of the clinical trial. It is essential that data analysis accurately reflects the research design to avoid bias in the results.

■ Some important considerations for the analysis of clinical trial data

Section V of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use states some important principles for clinical trial data management and data analysis [101]. The analysis datasets must be prospectively defined in the protocol to avoid selective inclusion of subjects in the analysis. The ‘all randomized patients’ should constitute the primary analysis dataset to ensure consistency with ‘intention-to-treat’ principle, which asserts that all randomized subjects are followed so long as alive, able, and consenting. The ‘per-protocol’ analysis set is a subset of all randomized patients who are compliant with their assigned treatment policy, have no major protocol violations, and have post-randomization measurements of the primary outcome. The per-protocol analysis plays a secondary role; however, ideally, it should yield consistent conclusions with the analysis based on all randomized subjects.

The intended data analysis strategies, including but not limited to the analyses of primary and key secondary variables, must be prespecified in the protocol. The choice of statistical methodology must be justified; strategies for checking model assumptions and alternative ways to analyze data if these assumptions are violated must be clearly spelled out in the protocol.

While, ideally, all patient data must be collected in a RCT, in reality some data will be missing due to patient dropouts, losses to follow up, and so on. Missing data

represents a potential source of bias in a RCT, because unobserved measurements may contain important information about treatment effects. Ways of handling missing data must be prospectively defined in the study protocol. General guidance on handling missing data in a RCT is available [102], but unfortunately there is no universal methodology for proper handling of missing data. It is generally acknowledged that the traditional popular Last Observation Carry Forward analysis should be avoided as the inference obtained using this methodology may be inefficient, biased, or both [58–60]. Some useful approaches for dealing with missing data include mixed-effects models for repeated measures for continuous outcomes, and generalized estimating equations and generalized linear mixed models for categorical responses and count data. For more details, see two recent special issues of the *Journal of Biopharmaceutical Statistics* devoted to missing data [61,62].

Finally, design considerations such as multiple comparisons and subgroup analyses must also be planned in advance to ensure credibility of the subsequent analysis.

■ The role of randomization in the analysis of clinical trial data

Perhaps the major strength of randomization in the context of a RCT is its ability to form the basis for statistical inference. Rosenberger and Lachin describe different models for statistical inference for the RCT [4]. The first one is the population model, which is built on the assumption that patients in the two treatment groups constitute random samples from infinitely large populations. This assumption is infeasible if one of the treatments is experimental – in that case there is simply no population to sample from. The invoked population model assumes there is an unspecified target patient population from which n eligible patients are sampled and then these patients are randomized to one of the treatment groups in the trial. The invoked population model assumes that patients have similar prognostic profiles, and errors in a statistical model are independent and identically distributed. In practice, this assumption is difficult to verify. However, randomization adds to the validity of this model as it introduces the real random element into allocation of treatments to study patients thereby making the assumption of independence of experimental errors more feasible compared with a nonrandomized study.

Unlike the invoked population model, the randomization model assumes that the statistical basis for inference is generated by randomization itself. This is built on RA Fisher's concept of randomization [62], which includes random allocation to treatments and generation of the probability space (reference set), by enumerating all possible randomization sequences of

the selected randomization procedure. The advantages of the randomization model are that it is completely nonparametric and it does not require that the study patients are homogeneous in terms of important prognostic factors that affect study outcomes. Statistical inference from randomization model guarantees a valid p-value even with very small sample sizes. The null hypothesis in the randomization model posits that the average treatment effect over all possible randomization realizations is zero. In contrast, a null hypothesis under a population model (Neyman–Pearson's) is based on the equality of parameters from known distributions. For testing randomization null hypothesis, one uses a randomization test. A statistic that measures the treatment group difference is chosen, and this statistic is computed for all possible realizations of the given randomization procedure; the sum of probabilities of those sequences whose test statistic values are at least as extreme as what was observed in the trial is the p-value of the test. A very small p-value indicates that there is strong evidence to conclude the difference between the treatment effects.

The randomization model for statistical inference has generated much controversy over the years; see [63,64] for interesting discussions on this matter. However, many authors hold the view that randomization-based inference is a useful alternative approach to population model-based methods [4,65] and it can be complementary to model-based inference. It is important to note that under randomization model, conclusions apply only to those patients who participated in the study and inferences on the broader population of similar patients will be non-statistical. The approaches to obtain generalizations in this case are description and analogy [66].

Finally, we note that although randomization-based inference is free of many population model-based assumptions and provides a useful tool to statistical inference, it is not immune to many operational biases we have described; therefore, it should not be viewed as a rescue tool for a poorly designed or conducted clinical trial.

Biases that randomization cannot handle

While randomization is helpful in mitigating experimental biases and contributes to the validity of statistical inference, it cannot eliminate biases that may arise due to the poor planning of a clinical trial. Incorrect assumptions on variability of the primary outcome measure may lead to a choice of sample size that is either too small or too large. Too small a sample size may result in a study that fails to detect a clinically meaningful treatment difference; too large a sample size can lead to a study that detects differences that are statistically significant but have no clinical relevance. The choice of the primary study objective, the research hypothesis,

the primary outcome measure and the sample size are all of great importance for the trial success. Even a very carefully implemented randomization design cannot alleviate issues that were overlooked at the planning stage. Likewise, randomization cannot handle biases related to interpretation and reporting of the results. This note reinforces the importance of all of the scientific, operational, and ethical aspects at all stages of a clinical trial.

Response adaptive randomization: useful or burdensome?

So far we have discussed randomization procedures that aim at achieving balance in treatment numbers (restricted randomization) or across covariates (stratified and covariate-adaptive randomization) and their ability to mitigate experimental biases. However, in clinical trials where primary outcomes follow nonlinear or heteroscedastic models, balanced allocation may not be optimal. Also, in clinical trials for fatal diseases there is a strong need to minimize exposure of patients to inefficient or toxic treatment arms and/or skew allocation towards the treatment group showing superior efficacy to maximize beneficial experience of the trial participants. These considerations call for randomization procedures for which treatment randomization probabilities for a given patient are modified based on the history of treatment assignments and responses (response-adaptive randomization) or based on the history of treatment assignments, responses, covariates and the covariate vector of the new patient (CARA randomization) to achieve prespecified, multiple experimental objectives (frequently expressed as a combination of statistical efficiency and ethical considerations) while maintaining the validity and integrity of the trial results. For instance, in a multiarm binary response trial the objective may be to minimize the expected number of treatment failures subject to minimal constraints on the treatment allocation proportions and the power of a homogeneity test [67]; in a survival trial where the variability and direction of the treatment effect differ for the patient groups within a treatment, the goal may be to skew allocation towards the treatment that is clinically best for a patient, while maintaining the power of a statistical test.

Theoretical aspects of response-adaptive and CARA randomization procedures are covered in the book by Hu and Rosenberger [5]. Both response-adaptive randomization and CARA randomization should be applicable in Phase II and III clinical trials. However, these methods have found little use in practice. Some recent papers call into doubt the usefulness of response-adaptive randomization [68,69]. The position of regulatory agencies on the use of response-adaptive randomization

is not encouraging either; see [103,104]. Rosenberger *et al.* point out that response-adaptive randomization procedures have established statistical properties and therefore these designs merit fresh consideration from clinical trialists [16]. Indeed, if the methods have well-established theory, and can be potentially advantageous over balanced randomization designs [67,70–73], why is there hesitation applying them in practice? What are the major concerns of the regulators? In this section we review experimental biases that may likely arise in clinical trials employing response-adaptive and CARA randomization and give our perspective on how to minimize potential for these biases.

■ Operational & procedural biases

Clinical trials employing response-adaptive and CARA randomization are much more complex than conventional trial designs. Therefore, strict processes for handling interim data must be in place. Both response-adaptive and CARA randomization designs require that data at interim points in the trial are unmasked for the purpose of estimation of treatment effects and modification of treatment randomization probabilities. Unmasking may introduce bias into the study. A fundamental prerequisite for maintaining integrity of the trial results is that management of the study should be no different before and after interim information are available [74]. Gallo discusses operational and procedural issues of adaptive designs, which apply to response-adaptive and CARA randomization as well [75,76]. He raises two important questions:

- (i) Who can have access to interim data for the purpose of interim analysis and implementing adaptation?
- (ii) What is the impact of knowledge of interim analysis results and inferences on trial integrity?

With regard to the first question, it is advised that adaptive trial designs utilize a trial logistic model that includes three parties: the sponsor; the external to the sponsor data monitoring committee charged with the review of interim safety data; and the independent statistical analysis center, consisting of subjects with necessary statistical expertise, charged with analysis of unmasked efficacy data and implementing design adaptations [77]. The separation of responsibilities for interim safety review (data monitoring committee) and interim efficacy review (independent statistical analysis center) should minimize potential conflict of interests [78].

With regard to the second question, it is essential that information about estimated treatment effects at interim analysis be kept as confidential as possible. Adaptation mechanisms can be described in general terms in the study protocol, but design parameters and

threshold values can be described in a separate document with more limited circulation. Study personnel performing design adaptations must be an authorized board of qualified individuals not otherwise participating in the trial. We feel, however, that since many modern response-adaptive and CARA randomization design methodologies are based on multiple considerations, it seems quite unlikely that an observer will be able to ‘reverse engineer’ the algorithm and determine the estimates that led to the observed change in the design.

Good practices and practical considerations for successful implementation of response-adaptive designs that are also applicable to response-adaptive and CARA randomization procedures can be found in [79] and [80].

■ Accrual bias

Rosenberger describes accrual bias, which may potentially arise in trials with response-adaptive randomization [81]. Patients may wish to be recruited later in the trial so as to benefit from the accumulated history of previous outcomes. This may cause bias if there is a difference in characteristics of patients recruited early and later in the trial. It is recommended that patients be masked to their sequence number in the trial [82]. Note, however, that accrual bias is irrelevant in trials dealing with emergency therapies.

■ Misclassification of responses

Both response-adaptive and CARA randomization rely on accumulating data from patients in the trial. Needless to say, these data must be of excellent quality to avoid bias in adaptations and in the analysis. Primary outcomes should be measured accurately and reliably. Tamura *et al.* report the results of a response-adaptive randomization, placebo-controlled clinical trial of the antidepressant fluoxetine, where a surrogate outcome measure was used to perform response-adaptive randomization [83]. In that trial, some subjects’ responses were misclassified and therefore some design adaptations were applied incorrectly. Li and Wang study impact of misclassification of binary responses on optimal allocation and show that misclassification may have detrimental effect on the validity of the results [84]. Special care must be taken to ensure high quality assessments and absence of errors in response-adaptive trials.

■ Statistical issues

Overall, response-adaptive and CARA randomization procedures have similar statistical properties to conventional randomization designs [16]. A minimum prerequisite for a statistically valid trial is the strong control of type I error [103,104]. Many modern response-adaptive randomization procedures meet this requirement asymptotically as they yield consistent and asymptotically normal estimators and treatment allocation proportions

[5]. Therefore, standard asymptotic tests and estimators can be applied to data generated from response-adaptive trials. However, for finite samples, simulation studies must be performed to ensure type I error is controlled under standard-to-worst-case experimental setups. The issues of delayed responses, population heterogeneity, statistical power and sequential monitoring of response-adaptive trials have been addressed theoretically, and are discussed in detail in Rosenberger *et al.* [16].

■ Open problems

Randomization-based tests have not been well studied for response-adaptive- and CARA-randomization procedures. Missing data issues are ubiquitous in clinical trials. While there are methods for dealing with missing data for conventional study designs [102], no such methods exist yet for response-adaptive and CARA randomization. Another important issue is related to modeling bias. CARA-randomization procedures rely on correctly specified parametric models for the primary outcome. If these assumptions are incorrect, estimated treatment effects both at the interim and in the final analysis will be biased and type I error may be inflated. There is no theoretical research into the effects of modeling bias on CARA-randomization designs, but simulations can be used to assess robustness of the designs to model misspecification.

Conclusion & future perspective

RCT carry the most credibility in the field of evidence-based medicine. While randomization can mitigate various potential experimental biases and contributes to statistical validity of the trial results, it must be complemented by other techniques such as masking, allocation concealment, stratification or adjustment of randomization for covariates (when appropriate) and careful conduct of the trial. All stages from the protocol development to the interpretation and dissemination of the clinical trial results are of great importance. Even the most scientifically sound randomization design will not provide protection from bias if the study is poorly planned or improperly conducted. Various randomization designs with well-established statistical properties are available for use in practice. The choice of the study design should be determined by the trial objectives, which should include statistical, ethical, and budgetary considerations. Some trials can be successfully implemented with a simple design using advance randomization or restricted randomization. Other trials may require more careful consideration of interplay among multiple study objectives and call for more special randomization designs, such as covariate-adaptive, response-adaptive or CARA. While the latter designs are more complex and require more rigorous planning than conventional randomization designs,

many of these designs have firm theoretical basis. Well-designed and well-documented simulation studies are necessary to understand operating characteristics of the designs under various experimental conditions [85,86]. Well-trained and qualified individuals with necessary medical and statistical expertise should be involved in

the implementation of such trials. Overall, we feel that modern advances in information technology, rigorous planning, and adherence to the key principles of the well-conducted clinical trial should enable successful implementation of response-adaptive and CARA randomization designs in the near future.

Executive summary

Randomization in clinical trials

- The randomized clinical trial is the most credible research design for clinical investigation. Various randomization procedures with established statistical properties are available for use in practice.

Bias & its control in randomized clinical trials

- To achieve high quality results, randomization must be complemented by other techniques, such as masking, allocation concealment, stratification or adjustment of randomization for covariates (when appropriate), covariate-adjusted analysis, and the careful conduct of the trial.
 - Ascertainment bias can arise in a clinical trial due to the knowledge of which treatment each participant is receiving. A randomized, double-masked trial design is the most efficient way to mitigate ascertainment bias.
 - Selection bias can arise due to an investigator's attempt to guess the treatment assignment and selectively enroll participants into the study. Randomization, masking and careful allocation concealment minimizes the potential for selection bias. In an unmasked trial, selection bias can be mitigated by the use of randomization procedures with minimal predictability. Selection bias can be also introduced at the data analysis stage by postrandomization exclusions. Such a bias can only be mitigated by sufficient attention and resources committed to active retention of study patients.
 - Covariate imbalances may arise by chance in clinical trials of any size. In a randomized trial with appropriate allocation concealment, covariate imbalances are random and do not compromise the validity of inference. An unbiased treatment comparison is achieved by a covariate-adjusted analysis.
 - Accidental bias refers to bias from unknown covariates that impact the primary outcome. Randomization tends to mitigate accidental bias; however, one should avoid using randomization procedures that can potentially generate a sequence of treatment assignments with some particular periodicity.
- Randomization cannot eliminate biases that may arise due to a poor planning of a clinical trial or biases related to interpretation and reporting of the trial results.

Special randomization designs: response-adaptive & covariate-adjusted response-adaptive

- Many modern clinical trials attempt to achieve multiple experimental objectives while maintaining the validity and integrity of the trial results. Such trials call for response-adaptive or covariate-adjusted response-adaptive randomization procedures. While these procedures have well-established theory, they are operationally more complex than traditional randomization designs and they have been rarely used in clinical trials.
- Modern advances in information technology, rigorous planning, and adherence to the key principles of the well-conducted clinical trial should enable successful implementation of response-adaptive and covariate-adjusted response-adaptive randomization procedures in the near future.

Acknowledgements

The authors thank four referees for the careful reading and comments that have substantially improved the earlier version of the paper.

Financial & competing interests disclosure

WF Rosenberger's research was supported by a grant from the National Science Foundation under the 2009 American Reinvestment and Recovery Act. The authors have no other 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 apart from those disclosed. No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as:

- of interest
 - ■ of considerable interest
- 1 Hadorn DC, Baker D, Hodges JS, Hicks N. Rating the quality of evidence for clinical practice guidelines. *J. Clin. Epidemiol.* 49(7), 749–754 (1996).
 - 2 Altman DG, Schulz KF, Moher D *et al.* The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann. Intern. Med.* 134(8), 663–694 (2001).
 - 3 Friedman LM, Furberg CD, DeMets DL. *Fundamentals of Clinical Trials (3rd Edition)*. Springer Science + Business Media (1998).
 - 4 Rosenberger WF, Lachin JM. *Randomization in Clinical Trials: Theory and Practice*. John Wiley and Sons, Inc., NY, USA (2002).
 - 5 Hu F, Rosenberger WF. *The Theory of Response – Adaptive Randomization in Clinical Trials*. John Wiley and Sons, Inc., NY, USA (2006).
 - 6 Chow SC, Liu JP. *Design and Analysis of Clinical Trials: Concepts and Methodologies (2nd Edition)*. John Wiley and Sons, Inc., NY, USA (2004).
 - 7 Piantadosi S. *Clinical Trials: A Methodologic Perspective (2nd Edition)*. John Wiley and Sons, Inc., NY, USA (2005).
 - 8 Gluud LL. Bias in clinical intervention research. *Am. J. Epidemiol.* 163(6), 493–501 (2006).

- 9 Jadad AR, Enkin MW. *Randomised Controlled Trials: Questions, Answers and Musings (2nd Edition)*. Blackwell Publishing, Malden, MA, USA (2007).
- 10 Liu Q, Chi GYH. Understanding the FDA guidance on adaptive designs: historical, legal, and statistical perspectives. *J. Biopharm. Stat.* 20(6), 1178–1219 (2010).
- 11 Pyke S, Julious SA, Day S *et al.* The potential for bias in reporting of industry-sponsored clinical trials. *Pharm. Stat.* 10(1), 74–79 (2011).
- 12 Berger VW, Bears JD. When can a clinical trial be called ‘randomized’? *Vaccine* 21(5–6), 468–472 (2003).
- 13 Berger VW, Ivanova A, Knoll MD. Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. *Stat. Med.* 22(19), 3017–3028 (2003).
- 14 Bailey R, Nelson PR. Hadamard randomization: a valid restriction of random permuted blocks. *Biometrical J.* 45(5), 554–560 (2003).
- 15 Efron B. Forcing a sequential experiment to be balanced. *Biometrika* 58(3), 403–417 (1971).
- **A seminal paper that introduced the biased-coin design and the concept of ‘accidental bias’.**
- 16 Rosenberger WF, Sverdlov O, Hu F. Adaptive randomization for clinical trials. *J. Biopharm. Stat.* 22(4), 719–736 (2012).
- 17 Day SJ, Altman DG. Blinding in clinical trials and other studies. *Brit. Med. J.* 321, 504 (2000).
- 18 Viera AJ, Bangdiwala SI. Eliminating bias in randomized controlled trials: importance of allocation concealment and masking. *Fam. Med.* 39(2), 132–137 (2007).
- 19 Altman DG, Schulz KF. Concealing treatment allocation in randomised trials. *Brit. Med. J.* 323, 446–447 (2001).
- 20 Forder PM, Gebiski VJ, Keech C. Allocation concealment and blinding: when ignorance is bliss. *Med. J. Aust.* 182(2), 87–89 (2005).
- 21 Berger VW. *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*. John Wiley and Sons, Inc., NY, USA (2005).
- 22 Berger VW, Weinstein S. Ensuring the comparability of treatment groups: is randomization enough? *Control. Clin. Trials* 25(5), 515–524 (2004).
- 23 Buyse M. Centralized treatment allocation in comparative clinical trials. *Appl. Clin. Trials* 9(6), 32–37 (2000).
- 24 Longford NT. Selection bias and treatment heterogeneity in clinical trials. *Stat. Med.* 18(12), 1467–1474 (1999).
- 25 Proschan M. Influence of selection bias on type I error rate under random permuted block designs. *Stat. Sinica* 4, 219–231 (1994).
- 26 Berger VW, Exner DV. Detecting selection bias in randomized clinical trials. *Control. Clin. Trials* 20(4), 319–327 (1999).
- 27 Markaryan T, Rosenberger WF. Exact properties of Efron’s biased coin randomization procedure. *Ann. Stat.* 38(3), 1546–1567 (2010).
- 28 Berger VW. The reverse propensity score to detect selection bias and correct for baseline imbalances. *Stat. Med.* 24(18), 2777–2787 (2005).
- 29 Ivanova A, Barrier RC, Berger VW. Adjusting for observable selection bias in block randomized trials. *Stat. Med.* 24(10), 1537–1546 (2005).
- 30 Blackwell D, Hodges JL. Design for the control of selection bias. *Ann. Math. Stat.* 28(2), 449–460 (1957).
- **Provides a mathematical model for selection bias and quantifies susceptibility of various restricted randomization procedures to selection bias.**
- 31 Deeks JJ, Dinnes J, D’Amico R *et al.* Evaluating non-randomised intervention studies. *Health Technol. Assess.* 7(27), 1–173 (2003).
- 32 Ioannidis JP, Haidich AB, Pappa M *et al.* Comparison of evidence of treatment effects in randomized and non-randomized studies. *JAMA* 286, 821–830 (2001).
- 33 Altman DG, Bland JM. Treatment allocation in controlled trials: why randomise? *Brit. Med. J.* 318, 1209 (1999).
- 34 Senn S. Testing for baseline balance in clinical trials. *Stat. Med.* 13(17), 1715–1726 (1994).
- 35 Senn S. *Statistical Issues in Drug Development (2nd Edition)*. John Wiley and Sons, Inc., NY, USA (2007).
- 36 Rosenberger WF, Sverdlov O. Handling covariates in the design of clinical trials. *Stat. Sci.* 23, 404–419 (2008).
- 37 Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. *J. Clin. Epidemiol.* 52(1), 19–26 (1999).
- 38 Hallstrom A, Davis K. Imbalance in treatment assignments in stratified blocked randomization. *Control. Clin. Trials* 9(4), 375–382 (1988).
- 39 Taves DR. Minimization: a new method of assigning patients to treatment and control groups. *Clin. Pharmacol. Ther.* 15(5), 443–453 (1974).
- **One of two papers (with [40]) that independently developed a covariate-adaptive randomization procedure (‘minimization’ method), as an alternative to stratified randomization.**
- 40 Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* 31(1), 103–115 (1975).
- **One of two papers (with [39]) that independently developed a covariate-adaptive randomization procedure (‘minimization’ method), as an alternative to stratified randomization.**
- 41 Atkinson AC. Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* 69(1), 61–67 (1982).
- **One of the first papers that introduced formal optimality in the design of a randomized comparative clinical trial.**
- 42 Taves DR. The use of minimization in clinical trials. *Contemp. Clin. Trials* 31(2), 180–184 (2010).
- 43 Pond GR, Tang PA, Welch SA, Chen EX. Trends in the application of dynamic allocation methods in multi-arm cancer clinical trials. *Clin. Trials* 7(3), 227–234 (2010).
- 44 Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Stat. Med.* 31(4), 328–340 (2012).
- 45 Senn S, Anisimov VV, Fedorov VV. Comparisons of minimization and Atkinson’s algorithm. *Stat. Med.* 29(7), 721–730 (2010).
- 46 McEntegart DJ. The pursuit of balance using stratified and dynamic randomization techniques: an overview. *Drug Inf. J.* 37, 293–308 (2003).
- 47 McEntegart DJ, Nicholls G, Byrom B. Blinded by science with adaptive designs. *Appl. Clin. Trials* 3, 56–64 (2007).
- 48 Brown S, Thorpe H, Hawkins K, Brown J. Minimization – reducing predictability for multi-centre trials whilst retaining balance within centre. *Stat. Med.* 24(24), 3715–3727 (2005).
- 49 Barbachano Y, Coad DS, Robinson DR. Predictability of designs which adjust for imbalances in prognostic factors. *J. Stat. Plan. Infer.* 138(3), 756–767 (2008).
- 50 Kuznetsova OM. On the second role of the random element in minimization. *Contemp. Clin. Trials* 31(6), 587–588 (2010).
- 51 Hasegawa T, Tango T. Permutation test following covariate-adaptive randomization in randomized controlled trials. *J. Biopharm. Stat.* 19(1), 106–119 (2009).

- 52 Shao J, Yu X, Zhong B. A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika* 97(2), 347–360 (2010).
- 53 Endo A, Nagatani F, Hamada C, Yoshimura I. Minimization method for balancing continuous prognostic variables between treatment and control groups using Kullback-Leibler divergence. *Contemp. Clin. Trials* 27(5), 420–431 (2006).
- 54 Stigsby B, Taves DR. Rank-minimization for balanced assignment of subjects in clinical trials. *Contemp. Clin. Trials* 31(2), 147–150 (2010).
- 55 Baldi Antognini A, Zagoraoui M. The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. *Biometrika* 98(3), 519–535 (2011).
- 56 Zhao W, Hill MD, Palesch Y. Minimal sufficient balance – a new strategy to balance baseline covariates and preserve randomness of treatment allocation. *Stat. Methods Med. Res.* doi:10.1177/0962280212436447 (2012) (Epub ahead of print).
- 57 Rosenberger WF, Rukhin AL. Bias properties and nonparametric inference for truncated binomial randomization. *J. Nonparametr. Stat.* 15(4–5), 455–465 (2003).
- 58 Wong WK, Boscardin WJ, Postlethwaite AE, Furst DE. Handling missing data issues in clinical trials for rheumatic diseases. *Contemp. Clin. Trials* 32(1), 1–9 (2011).
- 59 Mallinckrodt CH, Clark WS, David SR. Accounting for dropout bias using mixed-effects models. *J. Biopharm. Stat.* 11(1–2), 9–21 (2001).
- 60 Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharm. Stat.* 7(2), 93–106 (2008).
- 61 Special Issue: Missing Data – Prevention and Analysis. *J. Biopharmaceutical Stat.* 19(6), (2009).
- 62 Special Issue: Missing Data – Principal, Theory and Applications. *J. Biopharmaceutical Stat.* 21(2), (2011).
- 63 Basu D. Randomization analysis of experimental data. The Fisher randomization test (with discussion). *J. Am. Stat. Assoc.* 75, 575–595 (1980).
- Provides a thorough overview of the randomization-based inference using the Fisher randomization test.
- 64 Senn S. Controversies concerning randomization and additivity in clinical trials. *Stat. Med.* 23(24), 3729–3753 (2004).
- 65 Berger VW. Pros and cons of permutation tests. *Stat. Med.* 19(10), 1319–1328 (2000).
- 66 Finch PD. Description and analogy in the practice of statistics (with discussion). *Biometrika* 66(2), 195–208 (1979).
- 67 Tymofeyev Y, Rosenberger WF, Hu F. Implementing optimal allocation in sequential binary response experiments. *J. Am. Stat. Assoc.* 102, 224–234 (2007).
- 68 Korn EL, Freidlin B. Outcome-adaptive randomization: is it useful? *J. Clin. Oncol.* 29(6), 771–776 (2011).
- 69 Yuan Y, Yin G. On the usefulness of outcome-adaptive randomization. *J. Clin. Oncol.* 29(12), e390–e392 (2011).
- 70 Zhang L, Rosenberger WF. Response-adaptive randomization for clinical trials with continuous outcomes. *Biometrics* 62(2), 224–234 (2007).
- 71 Zhang L, Rosenberger WF. Response-adaptive randomization for survival trials: the parametric approach. *Appl. Stat.* 56(2), 153–165 (2007).
- 72 Sverdlow O, Tymofeyev Y, Wong WK. Optimal response-adaptive randomized designs for multi-armed survival trials. *Stat. Med.* 30(24), 2890–2910 (2011).
- 73 Cheung YK, Inoue LY, Wathen JK, Thall PF. Continuous Bayesian adaptive randomization based on event times with covariates. *Stat. Med.* 25(1), 55–70 (2006).
- 74 Golub HL. The need for more efficient trial designs. *Stat. Med.* 25(19), 3231–3235 (2006).
- 75 Gallo P. Operational challenges in adaptive design implementation. *Pharm. Stat.* 5(2), 119–124 (2006).
- 76 Gallo P. Confidentiality and trial integrity issues for adaptive designs. *Drug Inf. J.* 40, 445–450 (2006).
- 77 Benda N, Brannath W, Bretz F *et al.* Perspectives on the use of adaptive designs in clinical trials. Part II. Panel discussion. *J. Biopharm. Stat.* 20(6), 1098–1112 (2010).
- 78 Cook T, DeMets DL. Review of draft FDA adaptive design guidance. *J. Biopharm. Stat.* 20(6), 1132–1142 (2010).
- 79 Gaydos B, Anderson KM, Burnham N *et al.* Good practices for adaptive clinical trials in pharmaceutical product development. *Drug Inf. J.* 43(5), 539–556 (2009).
- 80 He W, Kuznetsova O, Harmer M *et al.* Practical considerations and strategies for executing adaptive clinical trials. *Drug Inf. J.* 46(2), 160–174 (2012).
- 81 Rosenberger WF. New directions in adaptive designs. *Stat. Sci.* 11, 137–149 (1996).
- 82 Rosenberger WF. Randomized play-the-winner clinical trials: review and recommendations. *Control. Clin. Trials* 20(4), 328–342 (1999).
- 83 Tamura R, Faries D, Andersen J, Heiligenstein J. A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *J. Am. Stat. Assoc.* 89, 768–776 (1994).
- 84 Li X, Wang X. Response-adaptive designs with misclassified responses. *Commun. Stat. A – Theor.* (2012) (In Press).
- 85 Benda N, Branson M, Maurer W, Friede T. Aspects of modernizing drug development using clinical scenario planning and evaluation. *Drug Inf. J.* 44(3), 299–315 (2010).
- 86 Smith MK, Marshall A. Importance of protocols for simulation studies in clinical drug development. *Stat. Methods Med. Res.* 20(6), 613–622 (2011).
- Websites
- 101 International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. Topic E9: Statistical principles for clinical trials (1998). www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf
- The International Conference on Harmonization E9 guidance outlines the key important considerations for the design, conduct and analysis of randomized clinical trials.
- 102 European Medicines Agency. Guideline on missing data in confirmatory clinical trials. EMEA/CPMP/EWP/1776/99 Rev. 1 (2010). www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf
- 103 Committee for Medicinal Products for Human Use. Reflection paper on methodological issues in confirmatory clinical trials with an adaptive design. CHMP/EWP/2459/02 (2007). www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003616.pdf
- 104 The US FDA. Guidance for Industry. Adaptive Design Clinical Trials for Drugs and Biologics (draft document) (2010). www.fda.gov/downloads/Drugs/./Guidances/ucm201790.pdf

