The Role of Randomization in Bayesian and Frequentist Design of Clinical Trial

Berchialla Paola[1], Gregori Dario[2], Baldi Ileana[2]

[1]Department of Clinical and Biological Sciences, University of Torino, Via Santena 5bis, 10126, Torino, Italy

[2]Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic and Vascular Sciences, University of Padova, Via Loredan 18, 35131, Padova, Italy

Aknowledgements

This research was supported by the University of Torino, Grant no. BERP_RILO_17_01.

Corresponding author

Paola Berchialla, PhD
Dept. of Clinical and Biological Sciences
University of Torino
Via Santena 5 bis
10126 Torino
e-mail: paola.berchialla@unito.it

Abstract

A key role in inference is played by randomization, which has been extensively used in clinical trials designs. Randomization is primarily intended to prevent the source of bias in treatment allocation by producing comparable groups.

In the frequentist framework of inference, randomization allows also for the use of probability theory to express the likelihood of chance as a source for the difference of end outcome. In the Bayesian framework, its role is more nuanced. The Bayesian analysis of clinical trials can afford a valid rationale for selective controls, pointing out a more limited role for randomization than it is generally accorded.

This paper is aimed to offer a view of randomization from the perspective of both frequentist and Bayesian statistics and discussing the role of randomization also in theoretical decision models.

Keywords. Clinical Trials; Bayesian Inference; Frequentist Inference; Randomization

Introduction

The origin of randomization in experimental design can be dated back to its application in psychophysics research in the late nineteenth century. However, systematic studies on the role of randomization began with Ronald Aylmer Fisher (1890–1962), who is credited with the promulgation of the randomization principle as pointed out by the philosopher Ian Hacking (Hacking 1988). In fact randomization was not widely recognized or accepted until Fisher applied it to agricultural research starting in the 1920's (Lee, Chen et al. 2012).

Much of the theoretical research related to randomization has been conducted in the domain of its application to clinical trials. The trial conducted by the (Medical Research Council 1948) to test the efficacy of streptomycin for the treatment of pulmonary tuberculosis is generally considered as the first application of randomization to clinical trials (Hall 2007).

Since then, the clinical trial with randomization (i.e. randomized trial) has increasingly gained popularity. In the late nineties they became the methodological cornerstone of the evidence-based medicine movement (Sackett, Rosenberg et al. 1996) and systematic reviews of randomized trials were placed at the apex of the evidence pyramid (Rosner 2012).

Fisher's requirement of randomization (1935) initiated a methodological revolution in experimental research. In his view randomization accomplished two tasks, one essentially qualitative and the other more technical: it eliminated bias and it enabled a valid basis for inference.

In a clinical trial, properly implemented randomization promotes comparability of treatment groups by eliminating the (selection) bias that can be introduced by selecting specific patients to receive a specific treatment. Randomization, however, brings a high probability of comparability with respect to unknown important covariates as well. Fisher's second reason for requiring randomization was that it provided a firm probability base for statistical inference relying on the randomization itself rather than on assumptions concerning the data generating process. Because treatments are assigned to patients using random number generators, the rules of probability could be invoked, and the likelihood that the result is not due to chance could be evaluated.

The first principle of randomization, "comparing like with like" is largely uncontroversial (Cox 2009). The possibility of bias in treatment allocation, in implementation, and in assessment of outcomes arises in many contexts in which subjectivity is even marginally involved and randomization is often the most effective way of avoiding such biases. Failure to randomize properly opens the door for confounding to walk in and may fatally compromise a clinical trial.

The second principle of randomization is a matter of question according to which statistical framework is chosen, whether frequentist of Bayesian. In a nutshell: Bayesian statistics is about making probability statements whereas frequentist statistics is about evaluating probability statements (Gelman 2008).

The role for randomization was challenged from a Bayesian point of view by Savage (Savage 1962) and several arguments have been put forward to criticize the role of randomization is necessary for scientific validity (Urbach 1993; Worrall 2007)

This paper aims to offer a view of randomization from the perspective of both frequentist and Bayesian statistics and discuss adaptive randomization as point of encounter for an ethical evaluation of clinical trials.

Randomization and inference

Frequentist

The concept of a population model most commonly underlies the development of a statistical test. In an experimental setting, this model assumes that the sample of patients is representative of a reference population and that the patients' responses for the outcome variable are independent and identically distributed (i.i.d.) from a distribution dependent on unknown population parameters. In the population model, n_E and n_C patients are randomly drawn from an infinite population of patients on treatment *E* and treatment *C*, respectively. Then patients' responses can be regarded as i.i.d. according to some probability distribution.

The likelihood function is central to the process of estimating the unknown population parameters as it measures the support provided by the data for each possible value of the parameter.

Under complete or restricted randomization, the treatment assignments are independent of patient responses and consequently of the population parameter, say θ, of interest (this will not be the case for response-adaptive randomization presented in the next section). Furthermore, responses depend only on the treatment assigned and are i.i.d. under a population model (Lachin 1988). Relying on these facts, it can be shown (Rosenberger and Lachin 2016) that the likelihood of the data after n patients randomized and evaluated for the outcome, L_n, reduces to $L_n \propto \prod_i L(y_i|t_i;\theta)$ where y_i and t_i are the realized treatment assignments and responses from the i-th patient.

Note that the likelihood is identical for any arbitrary sequence of treatment assignments, including non-random sequences.

Unfortunately, the assumption that patients are randomly drawn from an infinite population of patients on a given set of treatments (say, E and C) is hardly tenable in a clinical trial because patients are recruited from various sources by a non-random selection of sites and upon fulfilment of inclusion criteria and of requirement of informed consent. Therefore, the lack of a formal sampling basis does not justify the application of population models to clinical trials (Berger 2000).

Another approach to inference is based on the so-called randomization model that considers the probabilities of treatment assignment and their dependencies, if any.

The main difference between the randomization model and the population model lies in how they treat the outcome variable of interest and the assignment to treatment: in the former model the outcome is fixed and the treatment assignments (design points) are random; in the latter, the outcome variable is random at fixed values of the design points (Rosenberger and Lachin 2016).

Under the randomization model, assumption-free statistical tests for the equality of the treatments among the n patients actually enrolled, known as Fisher randomization tests (Fisher 1935), may be used.

The null hypothesis of a randomization test is that the assignment of treatment *E* versus *C* has no effect on the responses of the *n* patients. This probability statement is very different from a null hypothesis under a population model, which is typically about the equality of parameters from known distributions. Under the null hypothesis of randomization test, a patient's observed response is what would have been observed regardless of whether treatment *E* or *C* had been assigned. Then the observed difference between the treatment arms depends only on the way in which the *n* patients were randomized. Thus, given a randomization sequence and the associated responses for the outcome variable, a randomization test of the treatment effect can be built by permuting the sequence in all possible ways according to the randomization mechanism underlying it.

As stated in (Lachin 1988), statistical inference in a clinical trial must be viewed as a two-step process. The first step is to determine whether there is a difference between treatments and the randomization test provides an assumption-free test to answer this question. The second step is to ascertain the extent to which the observed results can be applied to the hypothetical population from which these patients arose. For this, it is necessary to invoke a population model. In this perspective, randomization-based inference may be seen as a useful alternative to, or complement to, traditional population model-based methods.

However, it worth mentioning that even if randomization minimizes unmeasured confounding when comparing treatment groups this could be not the case when analysing post-randomization factors, such as mediation variables. In fact, randomization in principle could not preclude potential confounding of the mediation relationships between a non-randomized factor and the outcome (Holland 1986).

Bayesian

In the Bayesian framework, the interest is in the inductive or subjective probabilities, which are computed through the Bayes Theorem.

For drawing inference, if the assignment mechanism, i.e. the mechanisms used to select experimental units and assign treatments, is not ignorable, then it must be modelled itself. The explicit inclusion of the assignment mechanism as random variable is pivotal in the Bayesian framework (Rubin 1991).

The distribution $g(T|X,Y,\theta)$ represents the probability of the treatment assignment given the data (X,Y) and it reflects the mechanisms that select the experimental units to assign to treatments. Thus $f(X,Y|\theta) \times g(T|X,Y,\theta)$ is the joint probability density function of the observations (X,Y,T) given θ and $p(\theta)$ is the conditional distribution of θ given the choice of the families of the models f and g.

For the sake of simplicity, here we consider the recording mechanism, i.e. the missing data-process, as ignorable. This is a choice justified by the fact that often unit labels, times of initiation of treatments and other aspects thought a priori to be uninteresting are not recorded. Such a priori decisions are completely specified, implying that the recording mechanism, i.e. which values are recorded for data analysis, is ignorable (Rubin 1978).

Since the conditional distribution of Y given X reflects a state of nature, the model $f(X,Y|\theta)$ is not under the researchers' control. However, the assignment mechanism in a context of clinical trials can be under the researchers' control, since they can assign treatments to the patients.

To be ignorable, the assignment mechanism must take the same known value for all the unknown X,Y,θ values (Rubin 1978). For example, in a sequential adaptive clinical trial, where the next patient receives the treatment based on past data, all values used in making the decision must be recorded. Similarly, in a randomized controlled clinical trial the distribution of the matching variables (i.e. age, sex etc.) must be recorded. On the contrary if the researchers assign the patients to a treatment according to their unrecorded clinical evaluation, or the patients select the treatment themselves, the assignment mechanism is not ignorable. Of course, the choice of the recording

mechanism can make the assignment mechanism not ignorable, except in the case of a simple random sampling followed by a randomized experiment, i.e. when $g(T|X,Y,\theta) = g(T|\theta)$.

Even if non-ignorable assignment mechanisms can be incorporated into the model leading to valid Bayesian inference, they pose greater problems then ignorable ones. In fact, it is common practice to choose vague prior distributions to reflect weak prior dependencies between parameters or to stretch out informative prior for computational reasons.

However, in general inference is very sensitive to prior specifications when the assignment mechanism is non-ignorable even with no imbalance in the distribution. Let's consider an example for explaining the role that randomization plays in the sensitivity of the posterior distribution. Consider two treatments *E* and *C* and consider the estimation of an average treatment effect in a small sample using a non-informative prior on the distribution of outcomes under each of the two treatments. The mechanism of assignment to each treatment has an effect on the variance of the posterior distribution. In fact, assigning randomly half of the subjects to each treatment is generally better than assigning each subject independently to the treatment, for example by flipping a coin. This happens because the expected posterior variance of the parameter of interest is smaller with balanced groups, i.e. equal number of subjects in each treatment group.

For example, compare the following two designs: assignment of the subjects to each treatment *E* or *C* at random or the systematic design *ECECECCECECE*. Both designs are ignorable given the covariates and thus a Bayesian inference of the outcome given the covariates is valid. However, the randomized design is ignorable even not given the covariates. Under the randomized design, the Bayesian analysis that pretends covariates are unknown is still valid, yielding a posterior predictive distribution $p(\tilde{y}|y)$. While pretending covariates are now observed, the posterior distribution can be updated to produce $p(\tilde{y}|y,x)$. Since both analyses are correct given their respective states of knowledge, they are expected to be consistent with each other with $p(\tilde{y}|y,x)$ more precise, i.e. smaller variance, than $p(\tilde{y}|y)$. If this is not, then the modelling assumptions should be reconsidered.

This extra step of model examination cannot be carried out under the systematic design without explicitly averaging over the distribution of the covariates. Thus, a randomized design increases the capability of performing posterior predictive checks (Gelman, Carlin et al. 2014).

Randomized allocation of patients to treatment has its own advantages, from a statistical point of view as well as from a clinical point of view since it allows for matching the trial groups. However, in the Bayesian framework the randomization, beyond being not necessary, is not even always the best mechanism of allocation of patients in the groups.

Adaptive randomization: point of encounter?

Clinical trials present a unique situation in which health-related interventions (i.e. drugs, biologics, etc.) are being tested for safety and efficacy. Until an intervention is proven to be effective and adequately safe, or ineffective or harmful, or just ineffective, the physician is in a state of equipoise: a state of genuine uncertainty about which experimental treatment is more effective (Hey and Truog 2015). In principle, it is ethical to employ randomization in a state of true equipoise. However, clinical trials, as experiments on humans, fed a heated debate on the ethics of randomization (Saxman 2015). The simplified key question is whether one should use equal randomization ratio (say 1:1 in case of two treatment groups) throughout the recruitment. Proponents of the application of equal randomization claim that equipoise should be retained until trial completion. Opponents advocate to vary the randomization ratio as trial data gets available. In other words, to specify randomization probabilities of treatment assignments conditional on the history of previous patients' treatment assignments, responses and/or covariates.

It is this ethical appeal that has motivated the research and application of adaptive randomization. Adaptive randomization alters the probability of patient allocation to different arms in order to meet a variety of objectives, while protecting the study from bias and preserving inferential validity of the results (Rosenberger 2010).

To motivate the use of adaptive randomization techniques in clinical practice, consider a two-arm clinical trial comparing an experimental treatment versus control. The most random procedure is the completely randomized design for which each subject is randomized between treatment arms with 0.5 probability and the assignments are mutually independent. This procedure balances treatment assignments asymptotically. However, it may likely result in large departures from balance in small samples (Cumberland and Royall 1988) and a randomization procedure which corrects such deviations is advisable. Furthermore, if accumulating trial data shows one treatment to be more promising, a randomization procedure which increases the probability of allocating patients to that treatment arm may be used.

In general, at the study start, all treatment arms have the same allocation ratio. As soon as information on outcomes is available, the randomization ratio can change to achieve a variety of objectives, including: *i)* correcting a chance deviation from the intended allocation ratio (i.e., restricted or treatment-adaptive randomization); *ii)* increasing the probability of assignment to the more effective or safe treatment (i.e., response-adaptive randomization); *iii)* balancing covariates (risk factors that modify the probability of an outcome) across different treatment arms (i.e., covariate-adaptive randomization). Some adaptive randomization procedures combine multiple methods such as the covariate-adjusted response-adaptive randomization.

Frequentist

We refer the reader to the book by (Rosenberger and Lachin 2016) for a technical survey of statistical methodologies for adaptive randomization.

Starting from the pioneering work of Efron's biased coin design (Efron 1971), restricted randomization methods, have been suggested in the literature in order to sequentially force the assignment toward balance by taking into account the history of previous allocations. At each step, the biased coin designs randomize the assignment by means of the tossing of a biased coin that

favours the treatment under-represented. The bias of the coin, *p*, represents a trade-off between balance and predictability.

(Wei and Durham 1978) were probably the first to discuss response-adaptive randomization. They proposed the so-called randomized play-the-winner rule which relies on the urn model and can be described as follows: at the outset, the figurative urn contains equal numbers of balls of each colour, with each colour associated to a different treatment group. Over time, the urn contains more and more balls with colours representing the arms with the more beneficial treatment, resulting in allocation of more patients to the most promising treatments.

This rule was used to design a paediatric trial of extracorporeal membrane oxygenation (ECMO), which compared the ECMO therapy versus the standard therapy (Bartlett, Roloff et al. 1985). Unfortunately, the trial provided very little information about survival rates of the two treatments. Out of 12 patients, only one was assigned to the standard therapy and died, whereas all 11 infants who were randomized to ECMO treatment survived. The main reason for the failure of the ECMO trial was the trial's small sample size and the poor operating characteristics of the play-the-winner rule, in particular, the rule's high variability and dependence on the initial composition of the balls in the urn (Rosenberger and Lachin 1993). Unfortunately, to this day, many investigators use the ECMO trial example as a reason to not to perform response-adaptive randomization at all.

The randomized play-the-winner is a heuristic procedure, while others exist based on optimal allocation targets. It must be noted that the adaptation of treatment allocation skews the treatment assignment and creates a dependency in collected data, thus reducing the statistical power to draw conclusions on treatment effect. Therefore, an optimal design of response-adaptive trials should consider both statistical power and the proportion of patients assigned to the more promising treatment. Several optimality criteria have been proposed to look for optimal allocations for response-adaptive designs in the consideration of statistical power. For example, the criterion proposed by Rosenberger, Stallard, Ivanova, Harper and Ricks (RSIHR) (2001) is to minimize the

expected number of treatment failures while keeping the conditional variance of the Wald test statistic at a fixed level.

The techniques for achieving covariate balance (i.e. covariate-adaptive randomization methods) encompass stratified (block) randomization, minimization, and dynamic hierarchical randomization (Lin, Zhu et al. 2015).

Bayesian

The natural capability of the Bayesian designs to deal with accumulated data as the trial progresses may be used for enabling dynamic allocation to experimental arms and dropping ineffective arms. This flexibility results in a potentially more efficient trial framework by increasing the probability of enrolment to arms that show evidence of efficacy.

Suppose the trial objective is to compare two treatments and patients are enrolled in sequential groups. Assume θ_i is the response rate, x_i is the number of responders, and n_i is the total number of patients for treatment i. Based on the standard binomial distribution, the distribution of the number of responders follows a Binomial distribution, and a conjugate beta prior distribution $P(\theta) = Beta(a, b)$ is usually chosen. In this scenario, the posterior distribution of θ_i is still a Beta distribution $P(\theta_i|x_i, n_i) = Beta\,(a + x_i, b + n_i - x_i)$ whose parameters are updated to the number of responders x_i and the number of non-responders $n_i - x_i$. A decision rule can be set to compare the response rate between the two treatments. For example, treatment 1 can be declared superior than treatment 2 if $P(\theta_1 > \theta_2) > 0.975$.

The standard study design randomizes with ratio 1:1 between the two treatments and compare the result at the end of study. On the other hand, an adaptive randomized procedure assumes that patients are enrolled over time and the interim results to can be used to preferentially allocate more patients into the more effective arm with a probability of allocation to treatment 1 given by

$$\frac{P(\theta_1 > \theta_2|data)^{\lambda}}{P(\theta_1 > \theta_2|data)^{\lambda} + P(\theta_2 > \theta_1|data)^{\lambda}}.$$

The case with $\lambda = 0$ corresponds to the scenario of equal randomization, whereas the case with $\lambda = \infty$ becomes the play-the-winner design (Berry, Bradley et al. 2010).

The Patient Assisted Intervention for Neuropathy: Comparison of Treatment in Real Life Situations (PAIN-CONTRoLS) is an illustrative trial of Bayesian adaptive randomization (Brown, Gajewski et al. 2016). PAIN-CONTRoLS trial is aimed at identifying which drug is most effective in reducing pain in patients with cryptogenic sensory polyneuropathy (CSPN). Pain at 12 weeks after study enrollment is the primary endpoint. Study participants are randomized to one of four drugs and pain is measured at 4, 8 and 12 weeks. Each participant at each measurement time is rated as keeping taking the drug or quitting due to lack of efficacy or adverse events.

After an initial phase where 80 participants are randomized with ratio 1:1:1:1 to the four arms (20 participants for arms), forward patients are allocated to one of the arms according to the posterior probabilities $(\theta_1, \theta_2, \theta_3, \theta_4)$ obtained from the data collected on the previous participants.

One way to increase the study efficiency is to incorporate stopping rules. Based on the interim result, if there is convincing evidence that one treatment is better than another, there is no need to continue the study, allowing better treatments to be adopted earlier.

In the PAIN-CONTRoLS trial, at each update at the 12-week visit a decision is made on whether the best drug is identified according to the rule that the posterior probability of the most effective drug is at least 0.925. Furthermore, any arm with posterior probability of no more than 0.01 is designated as loser with no more patients allocated to that arm. On the contrary, under the frequentist setting, the sequential stopping and the response-adaptive randomization are typically carried out separately because of complications that arise from dependence among the observations, which makes it difficult to justify the asymptotic properties of the design.

By carefully calibrating the design parameters, the loss of the statistical power due to groups imbalance can be controlled. At this stage, a key role in the adaptive randomization process is played in fact by the prior distribution.

A simulation carried out to assess the operating characteristics of the PAIN-CONTRoLS trial showed that in the scenario of one best drug, the type I error is approximately 5% and the power is about 94% with an average of 266 patients enrolled, below the 400 patients foreseen, half of them receiving the best treatment.

Discussion

Each method of randomization has properties that are better suited to specific applications than others. Thus, the choice of a randomization procedure and its implementation depend in part on the design features of the study.

Frequentism fits naturally with the regulatory "gate-keeping" role, through its insistence on procedures that perform well in the long run regardless of the true state of nature. And indeed, frequentist operating characteristics (Type I error and power) are key elements to regulators.

Although the Bayesian framework yield only incremental improvements over the frequentist's counterparts, it provides a uniform way of setting up complex problems, parameter estimation, and inference making. Bayesian framework also allows more flexible study conduct, such as dropping ineffective treatments and adding new treatments, because the inference is based on the data (conformed with the likelihood principle) and does not depend on a fixed sampling plan.

Randomization and balance are conflicting requirements and a suitable trade-off between optimality and predictability is crucial, for stopping a clinical trial at any time under an excellent inferential setting. In fact, while complete randomization prevents the selection bias of patients, it may led to imbalance distribution of some known prognostic covariates across groups.

On the other hand, treatment assignment based on the knowledge of patients' covariates can introduce a bias arising from the predictability of treatment allocation.

Covariate-adapting allocation procedure are considered an acceptable alternative to random assignment for achieving a compromise between balance and selection bias..

The Bayesian approach is a natural way to incorporate available data as a prior for decision making and therefore is advocated in response-adaptive randomization for clinical trials (Biswas, Liu et al. 2009). As acknowledged by (Thall and Wathen 2007) there is a large body of literature on adaptive randomization methods, both frequentist and Bayesian. Nevertheless, actual application of these methods to conduct clinical trials has been quite limited.

The adaptive randomization addresses the conflict between the common practice of using randomization versus the decision theoretic setup that would assign to the patients the optimal treatment. Utility based decision theoretic approaches to clinical trials provides a useful framework for ethical evaluation of clinical trials. A decision theoretic model is made of the space of all potential decisions, a utility (or loss) function and a probability model (Spiegelhalter, Abrams et al. 2004).

In this setting, the Bayesian inference provides a natural context to perform a formal decision theoretical approach, considering as probability model the posterior predictive model conditional on historical data.

An important aspect of fully decision theoretic approaches is the implication about randomization. Is a randomized decision justifiable for maximizing expected utility? As an extreme case, if a decision d_1 is equal in term of expected utility to decision d_2 then a random selection among them is justifiable, giving room to the frequentist inference, which does not require a model of outcomes but it does require a model of assignment (Christen, Muller et al. 2004; Lee et al. 2013).

However, such justification lacks the main motivations that leads researchers to choose randomization, which is to avoid biases arising from confounding and lurking variables. (Berry and Kadane 1997) proposed a formal justification of randomization by considering the impact of unknown covariates.

The relative merit of the Bayesian and frequentist approaches continues to be the subject of debate in statistics and other scientific fields. Regarding the two paradigms, we agree with (Lee and Chu 2012) that Bayesian and frequentist approaches offer complementary views and the future will be

cooperative at least on a practical level as recommended by (Little 2006) who suggests, based on the strengths and weaknesses of the two approaches, that inferences under a particular model should be Bayesian, but model assessment should involve frequentist ideas.

References

Bartlett RH, Roloff DW, Cornell RG, Andrews AF, Dillon PW and Zwischenberger JB (1985) Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study. Pediatrics **76**(4): 479-487.

Berger VW (2000) Pros and cons of permutation tests in clinical trials. Stat Med 19(10): 1319-1328.

Berry SM, Bradley P, Carlin J, Lee JJ and Muller P (2010) Bayesian Adaptive Methods for Clinical Trials, Chapman & Hall.

Berry SM and Kadane JB (1997) Optimal Bayesian Randomization Journal of the Royal Statistical Society. Series B (Methodological) 59(4): 813-819.

Biswas S, Liu DD, Lee JJ and Berry DA (2009) Bayesian clinical trials at the University of Texas M. D. Anderson Cancer Center. Clin Trials 6(3): 205-216.

Brown AR, Gajewski BJ, Aaronson LS, et al (2016) A Bayesian comparative effectiveness trial in action: developing a platform for multisite study adaptive randomization. Trials 17(1): 428.

Christen JA, Muller P, Wathen K and Wolf J (2004) Bayesian randomized clinical trials: A decision-theoretic sequential design. Canadian Journal of Statistics 32(4): 387-402.

Cox DR (2009) Randomization in the design of experiments. International Statistical Review 77: 415-429.

Cumberland WG and Royall RM (1988) Does simple random sampling provide adequate balance? Journal of Royal Statistical Society Series B 50: 118-124.

Efron B (1971) Forcing a sequential experiment to be balanced. Biometrika 58(3): 403-417.

Fisher RA (1935) The design of experiments. Oliver and Boyd, Edimburgh.

Gelman A (2008) Rejoinder. Bayesian Analysis 3(3): 467-478.

Gelman A, Carlin J, Stern H, Dunson D, Vehtari A and Rubin D (2014) Bayesian Data Analysis, Third Edition, Chapman and Hall/CRC, .

Hacking I (1988) Telepathy: Origins of Randomization in Experimental Design. Isis 79: 427-451.

Hall NS (2007) R. A. Fisher and his advocacy of randomization. J Hist Biol 40(2): 295-325.

Hey S P and Truog RD (2015) The Question of Clinical Equipoise and Patients' Best Interests. AMA J Ethics 17(12): 1108-1115.

Holland PW (1986) Statistics and Causal Inference. Journal of the American Statistical Association 81(396): 945-960.

Jiang F, Lee JJ and Muller P (2013) A Bayesian decision-theoretic sequential response-adaptive randomization design. Stat Med 32(12): 1975-1994.

Lachin JM (1988) Statistical properties of randomization in clinical trials. Control Clin Trials 9(4): 289-311.

Lee JJ, Chen N and Yin G (2012) Worth adapting? Revisiting the usefulness of outcome-adaptive randomization. Clin Cancer Res 18(17): 4498-4507.

Lee JJ and Chu CT (2012) Bayesian clinical trials in action. Stat Med 31(25): 2955-2972.

Lin Y, Zhu M and Su Z (2015) The pursuit of balance: An overview of covariate-adaptive randomization techniques in clinical trials. Contemp Clin Trials 45(Pt A): 21-25.

Little RJ (2006) Calibrated Bayes. The American Statistician 60(3): 213-223.

Medical Research Council (1948) STREPTOMYCIN treatment of pulmonary tuberculosis. Br Med J 2(4582): 769-782.
Rosenberger WF (2010) The agile approach to adaptive research. Wiley, New Jersey.
Rosenberger WF and Lachin JM (1993) The use of response-adaptive designs in clinical trials. Control Clin Trials 14(6): 471-484.
Rosenberger WF and Lachin JM (2016) Randomization in clinical trials: theory and practice. Wiley, New Jersey.
Rosenberger WF, Stallard N, Ivanova A, Harper CN and Ricks ML (2001). Optimal adaptive designs for binary response trials. Biometrics 57(3): 909-913.
Rosner AL (2012). Evidence-based medicine: revisiting the pyramid of priorities. J Bodyw Mov Ther 16(1): 42-49.
Rubin DB (1978) Bayesian Inference for Causal Effects: The Role of Randomization. The Annals of Statistics 6(1): 34-58.
Rubin DB (1991) Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. Biometrics 47(4): 1213-1234.
Sackett DL, Rosenberg WM, Gray JA, Haynes RB and Richardson WS (1996) Evidence based medicine: what it is and what it isn't. BMJ 312(7023): 71-72.
Savage LJ (1962) Subjective Probability and Statistical Practice, In: Savage LJ et al (ed) The Foundations of Statistical Inference. Methuen, London.
Saxman SB (2015) Ethical considerations for outcome-adaptive trial designs: a clinical researcher's perspective. Bioethics 29(2): 59-65.
Spiegelhalter D, Abrams K and Myles J (2004) Bayesian Approaches to Clinical Trials and Health Care Evaluation. Wiley, Chichester.
Thall PF and Wathen JK (2007) Practical Bayesian adaptive randomisation in clinical trials. Eur J Cancer 43(5): 859-866.
Urbach P (1993) The value of randomization and control in clinical trials. Stat Med 12(15-16): 1421-1431; discussion 1433-1441.
Wei IJ and Durham S (1978) The randomized play-the-winner rule in medical trials. J Am Stat Assoc 73: 840–843.
Worrall J (2007) Why There's No Cause to Randomize. The British Journal for the Philosophy of Science 58(3): 451-488.