CHAPTER 4

A Taxonomy of Classical Randomized Experiments

4.1 INTRODUCTION

In this chapter we introduce four specific examples of classical randomized assignment mechanisms, and we relate these examples to the general taxonomy of assignment mechanisms described in the previous chapter. The four examples, Bernoulli trials, completely randomized experiments, stratified randomized experiments (randomized blocks), and paired randomized experiments, all satisfy the four criteria necessary for assignment mechanisms to be classified as classical randomized experiments. These criteria, as discussed in more detail in Chapter 3, require that the assignment mechanism (*i*) is *individualistic*, with the dependence on values of covariates and potential outcomes for other units limited; (*ii*) is *probabilistic* – each experimental unit has a positive probability of being assigned to the active treatment and a positive probability of being assigned to the control treatment; (*iii*) is *unconfounded* – that is, given covariates, does not depend on potential outcomes; and (*iv*) has a known functional form that is controlled by the researcher.

The key difference between the four types of classical randomized experiments we consider in this chapter is in the set of assignment vectors $\mathbf{W}$ (the N-dimensional vector with elements $W_i \in \{0, 1\}$) with positive probability. Let the set of all possible values be denoted by $\mathbb{W} = \{0, 1\}^N$, with cardinality 2^N, and let the subset of values for $\mathbf{W}$ with positive probability be denoted by $\mathbb{W}^+$. In the first example of randomized experiments, Bernoulli trials, each of the 2^N possible vectors $\mathbf{W}$ defining the treatment assignments of the full population of size N has positive probability. However, such trials put positive probability on assignments in which all units receive the same treatment, thereby compromising our ability to draw credible and precise inferences regarding the causal effect of one treatment versus another from the resulting data. The remaining three types of classical randomized experiments impose increasingly restrictive sets of conditions on the set $\mathbb{W}^+$ of values of $\mathbf{W}$ with positive probability. If imposed judiciously, these restrictions can lead to more precise inferences by reducing the possibility of unhelpful assignment vectors (i.e., assignment vectors that *a priori* are unlikely to lead to useful inferences regarding the causal effects of interest).

https://doi.org/10.1017/CBO9781139025751.005 Published online by Cambridge University Press

4.2 NOTATION

In this section we briefly review the definition of, and notation for, classical randomized experiments, introduced in Chapter 3. The requirements for classical randomized experiments are that the assignment mechanism must be individualistic, probabilistic, and unconfounded and that the assignment mechanism is known to and controlled by the researcher. As a result of the first and third conditions, by Theorem 3.1, the assignment mechanism in a classical randomized experiment can be written as

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = c\cdot\prod_{i=1}^{N} e(X_i)^{W_i}\cdot(1-e(X_i))^{1-W_i},$$

for $\mathbf{W}\in\mathbb{W}^+$, and zero elsewhere. Here $\mathbb{W}^+\subset\mathbb{W}$ is the subset of the set of possible values for $\mathbf{W}$ with positive probability, and $e(x)$ is the propensity score, which, by probabilistic assignment, is strictly between zero and one. The constant c ensures that the probabilities add to unity:

$$c=\left(\sum_{\mathbf{W}\in\mathbb{W}^+}\prod_{i=1}^{N} e(X_i)^{W_i}\cdot(1-e(X_i))^{1-W_i}\right)^{-1}.$$

Because of the fourth condition, the propensity score $e(x)$ is a known function of the covariates. In this chapter we discuss four common classes of assignment mechanisms that fit into this framework: Bernoulli trials, completely randomized experiments, stratified randomized experiments, and pairwise randomized experiments.

4.3 BERNOULLI TRIALS

The simplest Bernoulli experiment tosses a fair coin for each unit: if the coin is heads, the unit is assigned the active treatment, and if it is tails, the unit is assigned the control treatment. Because the coin is fair, the unit-level probabilities and the propensity scores are all 0.5. Because the tosses are independent, the probability of any $\mathbf{W}$ for the N units in the study is the product of the individual probabilities; thus

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = 0.5^N, \tag{4.1}$$

for all $\mathbf{W}\in\mathbb{W}^+$. Here $\mathbb{W}^+=\{0,1\}^N=\mathbb{W}$.

Slightly more generally, we allow the probability of assignment to the treatment – that is, the propensity score – to be different from $1/2$, say $q\in(0,1)$. Then Equation (4.1) becomes

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = q^{N_{\rm t}}\cdot(1-q)^{N_{\rm c}}, \tag{4.2}$$

where $N_{\rm t}=\sum_{i=1}^{N}W_i$, and $N_{\rm c}=N-N_{\rm t}=\sum_{i=1}^{N}(1-W_i)$ are the number of treated and control units, respectively. Here, the probabilities of the different $\mathbf{W}$ vectors depend

https://doi.org/10.1017/CBO9781139025751.005 Published online by Cambridge University Press

solely on the number of treated and control units, but still $\mathbb{W}^+ = \{0, 1\}^N$. Such an assignment mechanism, where say, $q \in (0.5, 1)$, may be attractive, for example, when trying to induce people with a serious disease to enroll in a placebo-controlled experiment of a promising new drug for that disease. When the probability of assignment to the treatment group is higher than the probability of assignment to the control group, it would be more attractive for individuals to enroll in this trial than in one where the placebo or control treatment is as likely to be assigned as the active treatment.

Our final generalization of Bernoulli trials allows the unit probabilities to vary with the unit's covariate values. This situation can occur, for example, when certain types of patients are thought to do better on one treatment than another, and the strength of this belief about the better treatment varies with characteristics of the person (e.g., age, sex, race). Here, each unit has a special coin tossed, with the probability that the coin comes up heads equal to the probability that the unit is treated: the unit's propensity score. Consequently,

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{i=1}^{N} \left[e(X_i)^{W_i} \cdot (1 - e(X_i))^{1-W_i} \right]. \tag{4.3}$$

Here again $\mathbb{W}^+ = \mathbb{W}$. Our formal definition of a Bernoulli trial requires that assignments to treatment are independent across all units in the population:

Definition 4.1 (Bernoulli Trial)
A Bernoulli trial is a classical randomized experiment with an assignment mechanism such that the assignments for all units are independent.

Theorem 4.1 (Assignment Mechanism for a Bernoulli Trial)
If the assignment mechanism is a Bernoulli trial, then

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{i=1}^{N} \left[e(X_i)^{W_i} \cdot (1 - e(X_i))^{1-W_i} \right],$$

where $e(x)$ is the propensity score, which must be strictly between zero and one for all i, implying $\mathbb{W}^+ = \{0, 1\}^N$.

Proof. If assignment to treatment is independent across all observations in the population, then the probability of observing a specific assignment vector $\mathbf{W}$, $\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$, will simply equal the product of each unit's probability of assignment:

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{i=1}^{N} \left[p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))^{W_i} \cdot (1 - p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)))^{1-W_i} \right].$$

Combined with the fact that $p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = e(X_i)$ for all i, implied by the fact that a Bernoulli trial is a classical randomized experiment, it follows that the normalizing constant is $c = 1$ and that the general form of the assignment mechanism for this type of

https://doi.org/10.1017/CBO9781139025751.005 Published online by Cambridge University Press

randomized experiment is

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = \prod_{i=1}^{N}\left[e(X_i)^{W_i}\cdot(1-e(X_i))^{1-W_i}\right],$$

as in Equation (4.3). □

One common disadvantage of Bernoulli trials is that, because of the independence of the assignment across all units, there is always a positive probability (although small even in modest samples, and essentially zero in large samples) that all units will receive the same treatment. In that case, there will be no evidence in the data about the potential outcome values under the treatment that is not represented in the data. Even when there is a single unit being assigned one treatment and many assigned the other treatment, there will be limited evidence about the potential outcomes under the former treatment. Next, we therefore consider alternative classical randomized experiments that ensure that there are "enough" treated and control units under each assignment, beginning with the completely randomized experiment.

4.4 COMPLETELY RANDOMIZED EXPERIMENTS

In the second design we consider, the *completely randomized experiment*, a fixed number of subjects is assigned to receive the active treatment. The simplest completely randomized experiment takes an even number of units and divides them at random in two groups, with exactly one-half of the sample receiving the active treatment and the remaining units receiving the control treatment. This is accomplished, for example, by putting labels for the N units in an urn and drawing $N_{\rm t} = N/2$ at random to be treated. The assignment mechanism is:

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = \begin{cases} \binom{N}{N_{\rm t}}^{-1} & \text{if } \sum_{i=1}^{N} W_i = N_{\rm t}, \\ 0 & \text{otherwise,} \end{cases} \tag{4.4}$$

where

$$\binom{N}{N_{\rm t}} = \frac{N!}{N_{\rm t}!(N-N_{\rm t})!}.$$

The notation in (4.4) reveals that $N_{\rm t}$ does not have to equal $N/2$, but can be any positive integer less than N, fixed in advance. These designs are common in many applied settings, both because they assure that some units will be assigned each treatment, and because analyses using such designs are particularly straightforward in many circumstances. One reason for this simplicity is that the propensity scores are equal for all units, namely $N_{\rm t}/N$.

Definition 4.2 (Completely Randomized Experiment)
A completely randomized experiment is a classical randomized experiment with an assignment mechanism satisfying

https://doi.org/10.1017/CBO9781139025751.005 Published online by Cambridge University Press

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \,\middle|\, \sum_{i=1}^{N} W_i = N_\mathrm{t} \right\},$$

for some preset $N_\mathrm{t} \in \{1, 2, \ldots, N-1\}$.

In other words, given a population of size N, we fix the number of units assigned to the treatment, N_t such that $1 \leq N_\mathrm{t} \leq N-1$. Out of the population of N, we draw N_t units at random to receive the treatment. Each unit therefore has probability $q = N_\mathrm{t}/N$ of receiving the treatment. The number of possible assignment vectors, the cardinality of the set $\mathbb{W}^+$, is under this design $\binom{N}{N_\mathrm{t}}$. All $\binom{N}{N_\mathrm{t}}$ assignment vectors in $\mathbb{W}^+$ are equally likely; thus, the probability for any one is equal to $\binom{N}{N_\mathrm{t}}^{-1}$, whence in completely randomized experiments, the assignment mechanism is given by Equation (4.4).

Although often very sensible, completely randomized experiments are not without drawbacks, especially when important covariates are available. Important covariates here means covariates *a priori* thought to be possibly highly associated with the potential outcomes. Consider, for example, a study with $N = 20$ units, ten men and ten women, where the potential treatment and control outcomes are *a priori* thought to vary substantially by sex. Then, although a completely randomized design with $N_\mathrm{t} = 10$ would ensure that ten units get treated, there is the possibility that all ten of them are men (or women). In that case, average differences in the potential outcomes for active and control treatments could be due to sex differences rather than treatment effects. Related complications with relatively unhelpful (in the sense of being uninformative) experiments occur when only a single man is treated and nine men are in the control group, and so forth. The design studied in the next section addresses this issue in some circumstances.

4.5 STRATIFIED RANDOMIZED EXPERIMENTS

With the stratified randomized experiment, the population of units in the study is first partitioned into *blocks* or *strata* so that the units within each block are similar with respect to some (functions of) covariates thought to be predictive of potential outcomes. Then, within each block, we conduct a completely randomized experiment, with assignments independent across blocks.

The simplest randomized block experiment involves two blocks, say males and females, where independent completely randomized experiments are conducted for each group. There is no requirement that the numbers of males and females are the same. Thus, the assignment mechanism is the product of one expression like (4.4) for males, with $N(m)$ and $N_\mathrm{t}(m)$ replacing N and N_t, and one expression like (4.4) for women, with $N(f)$ and $N_\mathrm{t}(f)$ replacing N and N_t, with the experiment having a total of $N_\mathrm{t}(m) + N_\mathrm{t}(f)$ units assigned to the active treatment and has a total of $N(m) + N(f) - N_\mathrm{t}(m) - N_\mathrm{t}(f)$ units assigned to the control treatment.

In general, more strata can be used. Let $B_i \in \{1, \ldots, J\}$ indicate the block or stratum of the i^{th} unit, with $B_i = B(X_i)$ a function of the pre-treatment variables X_i, with a total

https://doi.org/10.1017/CBO9781139025751.005 Published online by Cambridge University Press

of J blocks or strata, and let $B_i(j)$ be the binary indicator for the event $B_i = j$. Then the assignment mechanism is the product of J versions of expression (4.4), each version having N and N_t indexed by the J distinct values of $B_i \in \{1, \ldots, J\}$. The unit-level probabilities are common for all units within a block but can vary across blocks. The main reason for generally preferring randomized blocks designs to completely randomized designs is that the former designs control balance in the covariates used to define blocks in treatment and control groups.

Formally, our definition of stratified randomized experiments is as follows:

Definition 4.3 (Stratified Randomized Experiment)
A stratified randomized experiment with J blocks is a classical randomized experiment with an assignment mechanism satisfying

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \,\middle|\, \sum_{i:B_i=j}^{N} W_i = N_t(j), \text{for } j = 1, 2, \ldots, J \right\},$$

and

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} \prod_{j=1}^{J} \binom{N(j)}{N_t(j)}^{-1} & \text{if } \mathbf{W} \in \mathbb{W}^+, \\ 0 & \text{otherwise,} \end{cases}$$

for some preset $N_t(j)$ such that $N(j) > N_t(j) > 0$, for $j = 1, \ldots, J$.

In this setting, the unit-level assignment probability or, equivalently in our situation with a classical randomized experiment, the propensity score, $e(X_i)$, is equal to $N_t(j)/N(j)$ for all units with $B_i = j$. As this representation makes explicit, this probability can vary with the stratum indicator. Often, however, the unit-level assignment probabilities are identical across the strata so that $e(x) = q$ for all x. In this case, the only difference between the stratified and completely randomized experiment is that in the former the relative sample size for treatment and control groups is constant across strata, whereas in the latter it may vary. If the covariates defining B_i correspond to substantive information about the units, in the sense that B_i is predictive of the potential outcomes, $(Y_i(0), Y_i(1))$, randomizing within the strata will lead to more precise inferences by eliminating the possibility that all or most units of a certain type, as defined by the blocks, are assigned to the same level of the treatment. Furthermore, even if there is no predictive power of the blocking indicator B_i, stratification does not reduce actual precision, though it reduces the number of allowable values of the assignment vector; see the notes to this chapter for some additional comments on this issue.

4.6 PAIRED RANDOMIZED EXPERIMENTS

The paired comparison, or randomized paired design, is an extreme version of the randomized block experiment in which there are exactly two units within each block, and a fair coin is tossed to decide which member of the pair gets the active treatment and

https://doi.org/10.1017/CBO9781139025751.005 Published online by Cambridge University Press

which gets the control treatment. As an example, consider an educational experiment with a covariate, a pre-test score, and the students are ranked from high to low on their scores on this pre-test. The top two form the first pair, the next two form the next pair, and so forth. Within each pair, one of the two units is randomly assigned to the treatment, with the probability of assignment equal to $1/2$.

Definition 4.4 (Paired Randomized Experiment)
A paired randomized experiment is a stratified randomized experiment with $N(j) = 2$ *and* $N_{\rm t}(j) = 1$ *for* $j = 1, \ldots, N/2$*, so that*

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \,\middle|\, \sum_{i:B_i=j}^{N} W_i = 1, \text{for } j = 1, 2, \ldots, N/2 \right\},$$

and

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} 2^{-N/2} & \text{if } \mathbf{W} \in \mathbb{W}^+, \\ 0 & \text{otherwise.} \end{cases}$$

In this design, each unit has probability $1/2$ of being assigned to the treatment group.

4.7 DISCUSSION

All four types of designs described in this chapter satisfy the four conditions for classical randomized experiments. In each case the assignment mechanism is individualistic, probabilistic, unconfounded, and known to the researcher. The way in which these four designs differ is in the set of values allowed for the vector of treatment indicators, $\mathbb{W}^+$. Reducing this set can be of great importance for the precision of estimated treatment effects. To illustrate this, consider the following example. Let N be even, and let the single pre-treatment variable X_i take on $N/2$ different values, with the number of units with $X_i = x$ equal to 2 for all $x \in \{1, \ldots, N/2\}$. Also assume identical unit-level assignment probabilities, that is, a constant propensity score, $e(x) = 1/2$ for all x. In Table 4.1 we report the number of values for the assignment vector that have positive probability under the various types of randomized experiments, for different sample sizes.

First, consider a Bernoulli trial. In this case, there are 2^N different values for the assignment vector. The first row in Table 4.1 shows that with $N = 4$ units, this corresponds to 16 assignment vectors. With $N = 16$, the number of possible treatment assignment combinations increases to more than 65,000.

Next consider a completely randomized experiment with $N_{\rm t} = N/2$ units assigned to treatment and $N_{\rm c} = N/2$ assigned to control. The number of allowed values for the assignment vector is now $\binom{N}{N/2}$, which is strictly less than the 2^N values allowed under the Bernoulli design. With $N = 4$ units, we now have only six possible assignment vectors; with a sample of $N = 16$, we have 12,870 possible assignment vectors, or roughly one-fifth the number possible with the Bernoulli trial.

https://doi.org/10.1017/CBO9781139025751.005 Published online by Cambridge University Press

Table 4.1. *Number of Possible Values for the Assignment Vector by Design and Sample Size*

Type of Experiment and Design	Number of Possible Assignments Cardinality of $\mathbb{W}^+$	Number of Units (N) in Sample			
		4	8	16	32
Bernoulli trial	2^N	16	256	65,536	4.2×10^9
Completely randomized experiment	$\binom{N}{N/2}$	6	70	12,870	0.6×10^9
Stratified randomized experiment	$\binom{N/2}{N/4}^2$	4	36	4,900	0.2×10^9
Paired randomized experiment	$2^{N/2}$	4	16	256	65,536

Third, consider a randomized block design, with two blocks, each consisting of $N/2$ units. Given our data set of N observations with the number of units with $X_i = x$ equal to 2 for all $x = 1, \ldots, N/2$, let the first block consist of all units with $X_i \leq N/4$, and the second block consist of the remainder. In terms of the notation introduced in Section 4.5,

$$B_i = \begin{cases} 1 & \text{if } X_i \leq N/4, \\ 2 & \text{if } X_i > N/4. \end{cases}$$

Suppose that within each block, the number of units assigned to the treatment group is equal to the number of units assigned to the control group, $N/4$. Now the number of values for the assignment vector within the first block is $\binom{N/2}{N/4}$, where this assignment vector $\mathbf{W}^{(1)}$ has $N/2$ components. In the second block the number of units is the same, $N/2$, so that the assignment vector for this block is also an $N/2$ component vector, $\mathbf{W}^{(2)}$, and the number of possible assignment vectors is again $\binom{N/2}{N/4}$. Therefore, the total number of values for the full assignment vector, $\mathbf{W} = (\mathbf{W}^{(1)}, \mathbf{W}^{(2)})$, possible under this design is the product of the within-block number of possibilities, $\binom{N/2}{N/4}^2$. Note that this is a strict subset of the set of possible values under the previous two designs. With $N = 4$ units, we now have only 4 possible assignment vectors; with a sample of 16, the number of possible assignment vectors is 4,900.

Fourth, consider the paired randomized experiment where units with the same value of X_i are paired, so $B_i = X_i$. Now there will be $2^{N/2}$ different possible values of the assignment vector with positive probability. This design is a randomized block experiment in which each stratum (block, or subclass) contains only two units. This assignment mechanism is also a paired randomized experiment. Note also that in a paired randomized experiment, using the same argument as above, any value of the assignment vector with positive probability under this design also has positive probability under the stratified randomized design. With only 4 units, the number of assignment vectors with positive probability under a paired randomized experiment is, in fact, identical to that with possible probability under a stratified randomized experiment. With only $N = 4$ units, in the

https://doi.org/10.1017/CBO9781139025751.005 Published online by Cambridge University Press

stratified design there can be at most 2 strata, each with the 2 units of a pair, and within each, only one observation assigned to the treatment. With 16 units, however, under a paired randomized experiment there are 256 assignment vectors with positive probability, compared to the 4,900 with positive probability under a randomized block design with two blocks, or a total of 65,536 values for the assignment vector with positive probability under the Bernoulli design.

In this particular sequence of designs with fixed N, the number of distinct values of the assignment vector with positive probability, that is, the cardinality of the set $\mathbb{W}^+$, gradually decreases. The argument for choosing successively more restrictive designs is to eliminate "unhelpful" assignment vectors that are *a priori* unlikely to lead to precise causal inferences. Imposing the first restriction – from Bernoulli trials to completely randomized experiments – is obvious. An assignment vector with all, or almost all, units assigned to one of the treatment levels is typically not as informative as an assignment vector with more balance between the number of treated and control units. Hence, a completely randomized design will tend to be more informative than a Bernoulli trial. The further restrictions to stratified and paired randomized experiments have similar advantages, when the grouping into strata or pairs is based on covariates that are related to the potential outcomes. Formally, if the information used in defining the blocks or pairs is relevant for predicting the potential outcomes, $(Y_i(0), Y_i(1))$, then these designs can improve on completely randomized experiments in terms of the precision of the estimates obtained, often considerably so. In an extreme case, if the pre-treatment variable, X_i, upon which the stratification or pairing is based, perfectly predicts both potential outcomes, there will be no uncertainty remaining regarding the treatment effect across the N units or within the subgroups defined by the covariate. On the other hand, if the blocks or pairs are formed in a way unrelated to the potential outcomes (e.g., by randomly drawing units to assign block labels B_i), the eliminated assignment vectors are just as likely to be helpful as the retained ones, and in such cases, the precision of estimators for treatment effects in stratified or paired randomized experiments is usually no greater than that for the corresponding estimators under completely randomized experiments.

In the next chapters, we discuss analyzing results from the various types of classical randomized experiments in more detail and illustrate these analyses with real data. The methods for analyzing these randomized experiments are useful for two major reasons. First, they are valuable in their own right for analyzing randomized experiments. For many questions in the biomedical and social sciences, however, we must rely on data from observational studies. The second use of these methods, as templates for the analysis of data from observational studies, are therefore even more important for us. In Parts III through VI of this text, we extend these methods for analyzing specific types of classical randomized experiments to assessing data from observational studies and show that observational data can often be analyzed as if they fit the assumptions of one of the randomized experiments discussed here.

4.8 CONCLUSION

In this chapter we discuss four special cases of classical randomized experiments: Bernoulli trials, completely randomized experiments, stratified randomized experiments,

https://doi.org/10.1017/CBO9781139025751.005 Published online by Cambridge University Press

and paired randomized experiments. In the next seven chapters we discuss and illustrate methods for estimation and inference in these settings. This is important for substantive reasons but also because understanding the analysis of such relatively simple cases is important for analyzing the more complex observational studies that are the subject of Parts III through VI of this text.

NOTES

There is a large classical literature on experimental design and the analyses of randomized experiments, including Cochran and Cox (1957), Cox (1958), Kempthorne (1952), and Box, Hunter, and Hunter (2005). Much of the design literature focuses on the optimal design of more complex studies with multiple treatments. Such questions are beyond the scope of the current text. Rosenbaum (2000) discusses the structure of the set of assignment vectors using results for finite distributive lattices. Morgan and Rubin (2012) discuss an additional class of designs for randomized experiments. The idea is to start with a completely randomized design. Then, given the assignments, balance of the covariates is assessed according to some well-defined criterion, articulated *prior* to the randomization. If the balance is deemed inadequate, the assignment is rejected and a new vector of assignments is drawn. This is repeated until an assignment vector is drawn that is deemed adequately balanced. Such designs can lead to more precise inferences than completely randomized designs, and they can be more attractive than stratification in settings with many covariates. A similar but different design is described by Morris (1979).

For general discussions of the literature on analyses of randomized experiments, see Altman (1991), Wu and Hamada (2009), Cook and DeMets (2008), Davies (1954), Cox (1958), Cochran and Cox (1957), Kempthorne (1957), and Box, Hunter, and Hunter (2005).

Imbens (2011) analyzes the gains from the stratification and shows that even in the absence of any dependence between the potential outcomes and the stratum indicators, stratification, in expectation, in settings with random draws from large strata, does not increase the actual sampling variance of simple estimators of the average treatment effect, thus showing that there is no cost in expected precision of estimation when using stratification even when the samples drawn from the strata are small. There are, however, fewer "degrees of freedom" to estimate that precision, and so the resulting inference is somewhat less precise, an issue studied first in Fisher (1935, pp. 248–250) from a fiducial-likelihood perspective. Specifically, Fisher suggests using the expected information, that is, the expected second derivative of the log-likelihood to adjust for this effect by multiplying the estimated sampling variances by $(K + 3)/(K + 1)$, where K is the number of degrees of freedom used to estimate each sampling variance. It is important here that the strata are large. If the strata are small in the population, it is possible that outcomes within strata are negatively correlated. Snedecor and Cochran (1967, p. 294) discuss examples where this may be relevant (e.g., rats' weights within a litter).