7

RANDOM ASSIGNMENT

Just as representativeness can be secured by the method of chance . . . so equivalence may be secured by chance.[1]

—W. A. McCall

LEARNING OBJECTIVES

- Understand what random assignment does and how it works.
- Produce a valid randomization process for an experiment and describe it.
- Critique simple random assignment, blocking, matched pairs, and stratified random assignment.
- Explain the importance of counterbalancing.
- Describe a Latin square design.

Just as the mantra in real estate is "location, location, location," the motto in experimental design is "random assignment, random assignment, random assignment." This book has discussed random assignment all throughout. It bears repeating that random assignment is the *single* most important thing a researcher can do in an experiment. Everything else pales in comparison to having done this correctly.[2] Random assignment is what distinguishes a true experiment from a quasi, natural, or pre-experimental design. In chapter 1, experiments were referred to as the gold standard. Without successful random assignment, however, they can quickly become "the bronze standard."[3] This chapter will review some of the advantages of random assignment, discuss the details of how to do it, and explore related issues of counterbalancing.

THE PURPOSE OF RANDOM ASSIGNMENT

People vary. That is, they are different. Were this not so, there would be no reason to study them. Everyone would be the same, reacting the same way to different teaching techniques, advertisements, health interventions, and political messages. There would be no need to conduct experiments. The fact that people vary provides social scientists with a reason for doing research, and also with their biggest challenge. When people vary on things that are not of interest to the experiment, it is called **random variation** or **noise**.[4] These things that are not the focus of a study can still be responsible for some of the changes in the outcome of experimental treatments. That is, they can confound the results. One major focus of experimentalists is to control for confounds, removing or reducing the noise from random variation. That way, the effects the study is concerned with can be seen more clearly. Randomization is arguably the greatest weapon a scientist has because it helps ensure that subjects in different treatment and control groups are virtually the same on variables that create noise.

Random assignment is a technique for placing subjects into the different treatment and control groups in a true experiment for the purpose of ensuring that subjects in each group will have similar characteristics—that is, that they will be **equivalent**. By equivalent, we mean equal on average, or probabilistically equal, not identical. The purpose of equivalence is to "level the playing field," helping ensure that the only systematic differences between the two groups are the treatments they receive in the experiment. It helps ensure that subjects in one group are not better on the outcome variable to begin with, and that one group is not filled with subjects who are more likely to change regardless of treatment.[5] This allows researchers to have more confidence that any changes observed are because of the treatment the subjects received and not because of inherent differences in the subjects themselves. Groups that are systematically different in one or more ways can invalidate an experiment. For example, if one group contained only men and the other only women, it would be confounded; there would be no way to tell if the results were due to the treatment or to gender. One way random assignment helps achieve equivalence is by avoiding **selection bias,** which occurs when people self-select the groups to be in, or researchers select subjects for some subjective reason or to improve the chances of supporting a hypothesis, even if done subconsciously.[6]

For example, an education study would be invalid if the treatment group gets mostly subjects who are better at math to begin with and the control group gets subjects who are poor at math. If the treatment is the way math is taught, then it might appear as if the treatment is working, but it really could be the fact that those who were taught the new way were better at math to begin with. On the other hand, if students who are poor at math are assigned to the treatment group by educators who want to be sure that students who need math help the most get it, then the treatment group is stacked with poor-math students and the control group with good-math students. The new teaching strategy could actually work, but might look like it was not if it only raised the treatment group up to the level of the controls; in other words, there is no significant difference between the groups after treatment.

Another example is that some people are simply more prone to change than others. If the treatment group got more subjects who changed more easily than the control group, then a study designed to influence people's positions on public policy could show spurious results—the treatment did not really change people's minds; they were more likely to change their minds to begin with. Giving the same treatment to harder-to-change people might have no effect at all.

Avoiding Confounds

The beauty of random assignment is that the researcher does not have to predetermine every characteristic of every subject that could possibly confound the study. It might be easy to anticipate that prior math knowledge would influence math learning, for example, and pretest subjects on math and assign equivalent numbers of high-math knowledge and low-math knowledge subjects to each group. But it might be more difficult for a researcher to anticipate a penchant for changing one's mind as a confounding variable. With random assignment, no pretests for math ability or mind changing are necessary because equivalent numbers of easy-changers and hard-changers are in each group; or equivalent numbers of good-at-math and poor-at-math students are in each group. Because humans, including researchers, are notoriously bad at anticipating every little thing that might affect something else, and because some things are unknowable, random assignment is of tremendous benefit. Also, recall the drawbacks of pretesting from chapter 4; random assignment can eliminate the need for pretests and their accompanying threats to validity.

Random assignment, while not perfect, is the best way we currently know of to ensure that systematic variation among subjects does not confound the results of an experiment. It helps eliminate spurious variables, including those a researcher might not have thought of.

When it is impossible for the same people to simultaneously have the treatment and not have it, covered in chapter 6, researchers use random assignment to try to make sure that individual differences are assigned equivalently to each group.

What Is Random?

Random assignment works because of chance. To assign subjects randomly, everyone in the study must have an equal chance of being in the control or treatment groups. Chapter 2 recounted the story of how random assignment was discovered. When Charles Sanders Peirce and Joseph Jastrow conducted a study of how accurately people could judge the weight of something just by feeling and looking at it,[7] they started by presenting the heavy weights first. Then they presented the heavy weights last. They also alternated the heavy and light weights. Finally, they shuffled a deck of cards and assigned the weights at random based on the cards. They got vastly different results when the weights were presented in any of the systematic patterns than when they were presented in a random order. Having subjects who could not guess the weight based on a pattern produced more valid results.

Random does *not* mean haphazard, and researchers must be careful to use appropriate random methods.[8] What is not random is anything that has some kind of pattern, purpose, or system to it. In my first experiment, I apparently did not understand exactly what random meant, but I was concerned with having equal numbers of subjects in each group. I assigned the first person to come into the lab to the first group, the second person to the second group, then the third person back to the first group, etc. Basically, they were assigned like this: 1, 2, 1, 2, 1, 2, 1, 2 . . . . When I told my professor, her eyes got wide and she said, “That’s NOT random.” I ended up throwing away all that data and starting over again.

As Gueron says, “It does not help to be a little bit random.”[9]

Dilbert © 2001 Scott Adams. Used By permission of Andrews McMeel Syndication. All rights reserved.

MORE ABOUT . . . BOX 7.1

Random Assignment Threats to Internal Validity

When subjects and/or the experimenters that assign subjects to condition are not "blind" to the group they are being assigned to, random assignment itself, and violations of it, can result in threats to internal validity. The four types are:[13]

- *Diffusion of treatment*—This occurs when subjects in different groups communicate with each other and learn information not intended for them. For example, if subjects in the treatment and control group talk to each other, they may learn material intended only for the treatment group; the outcomes may not be different and will not truly reflect the treatment benefits.
- *Compensatory equalization*—This occurs when experimenters or those providing the treatment attempt to give some of the advantages that the treatment group has to those in the control group.
- *Compensatory rivalry*—This refers to when members of the control group try to gain some of the benefits of the treatment group.
- *Resentful demoralization*—This refers to the control group subjects underperforming because they resent being denied the treatment.

In some studies, subjects in the treatment group feel demoralized or stigmatized—for example, as being in the class for "dummies."[14] Thus, it is important that subjects be blind to which group they are in whenever possible.[15]

Researchers should be careful not to portray one intervention group as better or newer than another in recruitment materials.[16] If it is not possible to keep subjects from knowing which group they are in, researchers should measure subjects' preferences for assignment to a particular group and statistically control for it.[17]

Another threat to internal validity occurs when there is an imbalance of subjects who have a greater preference for their assignment in one group than in the other.[18] For example, if 60% of subjects in both groups are pleased with the group they are in and 40% are not, then preference is equivalent across all conditions. However, if one group has 60% pleased and the other has only 40% pleased, the outcomes could be confounded.[19] As long as both groups have the same percentage of pleased and displeased subjects, no threat occurs.[20] For example, in a business study, 89% of those asked to be in the treatment group agreed to participate, whereas only 45% of those asked to be in the control group agreed.[21]

True randomization is a process that is either done correctly or not.[10] In order for random assignment to work, the researcher cannot choose which group a subject is in for any reason, nor can a subject choose his or her own group. The subjects must be **blind,** or unaware, as to whether they are receiving the treatment or not. It is important that subjects remain in the dark so they do not try to give researchers what they think they want,[11] or so their disappointment at not getting the treatment does not bias the results.[12] It is also a good idea to have the researcher be unaware of which treatment or control groups subjects are in, which is then termed a *double-blind* study. For more about threats to internal validity when subjects and/or researchers know which group they are assigned to, see More About box 7.1.

OPERATIONALIZING RANDOM ASSIGNMENT

There are many ways to "do" random assignment. In 1883, Peirce and Jastrow used a special deck of cards to determine random assignment, and that is still a valid method today. Other methods include rolling dice, flipping a coin, drawing numbers out of a hat, or using a book of random numbers. For details on how to do these manual methods, see How To Do It box 7.2.

Computerized Randomization

Today, it is more likely that researchers will use random number generators found online or in spreadsheet or statistical software. For example, free online randomizers allow a researcher to specify the number of subjects and number of groups, and quickly return a list showing which subjects go to which groups, as shown in figures 7.1 and 7.2 (see pp. 180–181). Simply search the Internet for "random assignment generator" to find these.

To use QuickCalcs by GraphPad, in the first box labeled "Assign," put in the total number of subjects. Put the number of groups in the second box. Leave "Repeat" at 1. Click "Do it!"

It will return a list of subjects, numbered 1 to your final number, with the group labeled A, B, C, etc., beside the subject number.

This applet is specifically designed by a group of academic researchers for factorial experiments. In the first box, "Number of Participants, *N*" put the total number of subjects for the entire study. In the second box, "Number of Conditions, *C*," put the number of groups. Click "Compute."

It will return a list of groups to assign participants to, in order; it lacks the subject number of the QuickCalcs randomizer, but it is not hard to see that the first subject is assigned to the first group number, the second subject to the group number listed next, etc.

HOW TO DO IT 7.2

Randomizing Subjects

Coin flipping works when there are two groups. If the coin lands heads up, assign the subject to the treatment group; if it lands tails up, assign the subject to the control group. Or vice versa.

A *lottery* works with any number of groups. On slips of paper, write a number for each of the groups (1, 2, 3, 4, for example). Make as many slips of paper as subjects you intend to have, with equal amounts of group numbers. For example, if you need 160 subjects, with forty in each group, make forty slips of paper with the number 1, forty with number 2, etc.

Decks of cards and dice will need to use cards and die with only the number of groups on them. For example, get rid of all cards that are not 1, 2, 3, or 4 if you have four groups. You might need multiple decks. Shuffle the cards and then draw them one at a time each time a subject arrives. The subject is assigned to the group represented on the card drawn. For dice, use only dice that have the same number on them as your groups. Or roll again if a number comes up that is greater than your number of groups. Once the maximum number of subjects in a group is reached, ignore that number when it comes up on a card or die.

Book of random numbers. These are obsolete today, replaced by online random number generators. They consisted of page after page of numbers, randomly ordered. Believe it or not, a researcher decided where to start by closing his or her eyes and pointing to a starting place on the table. Then, the researcher would assign subjects to the group that the numbers correspond to, skipping over numbers that are outside the range. For example, if the numbers are: 3, 2, 5, 3, 7, 4, 1, etc., and the study has three groups, subjects would be assigned to groups 3, 2, 3, 1, etc., skipping over 5 and 7 because there are not five or seven groups. Once the maximum number of subjects in a group has been reached, that group's number is skipped over as well. Stop assigning subjects to a group after a group reaches the maximum, but continue on until all the other groups have been filled. My favorite book of random numbers was *A Million Random Digits With 100,000 Normal Deviates* (RAND). Your library might still have it if you are curious or a history buff.

Excel. For a tutorial on how to draw random numbers in Excel, see https://exceljet.net/formula/randomly-assign-data-to-groups.

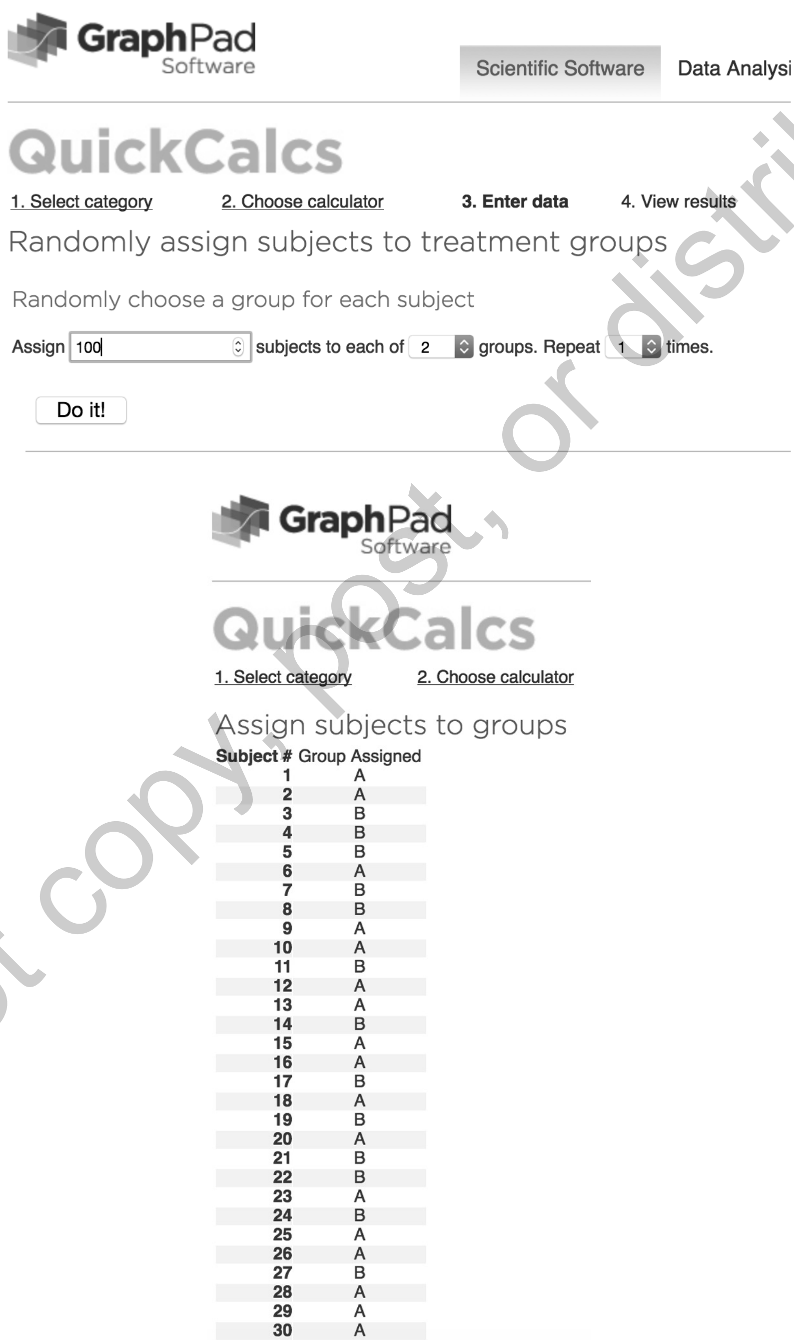
SAS and SPSS. Shadish, Cook, and Campbell[22] give directions on random number generation in these popular statistical software packages on pages 311–313.

For other statistical software, such as *Stata* and *R*, consult the tutorials.

For pencil-and-paper studies, an online randomizer can be used to arrange the printed questionnaires in the specified random order before going to the lab or site where subjects will take the study. Questionnaires can then be handed out from the top of the stack without needing to refer to the output of random numbers each time a subject arrives.[i]

[i]Be sure to make a notation on the paper questionnaire or use another method to keep track of which group, treatment, or control the subject was in if it is not obvious.

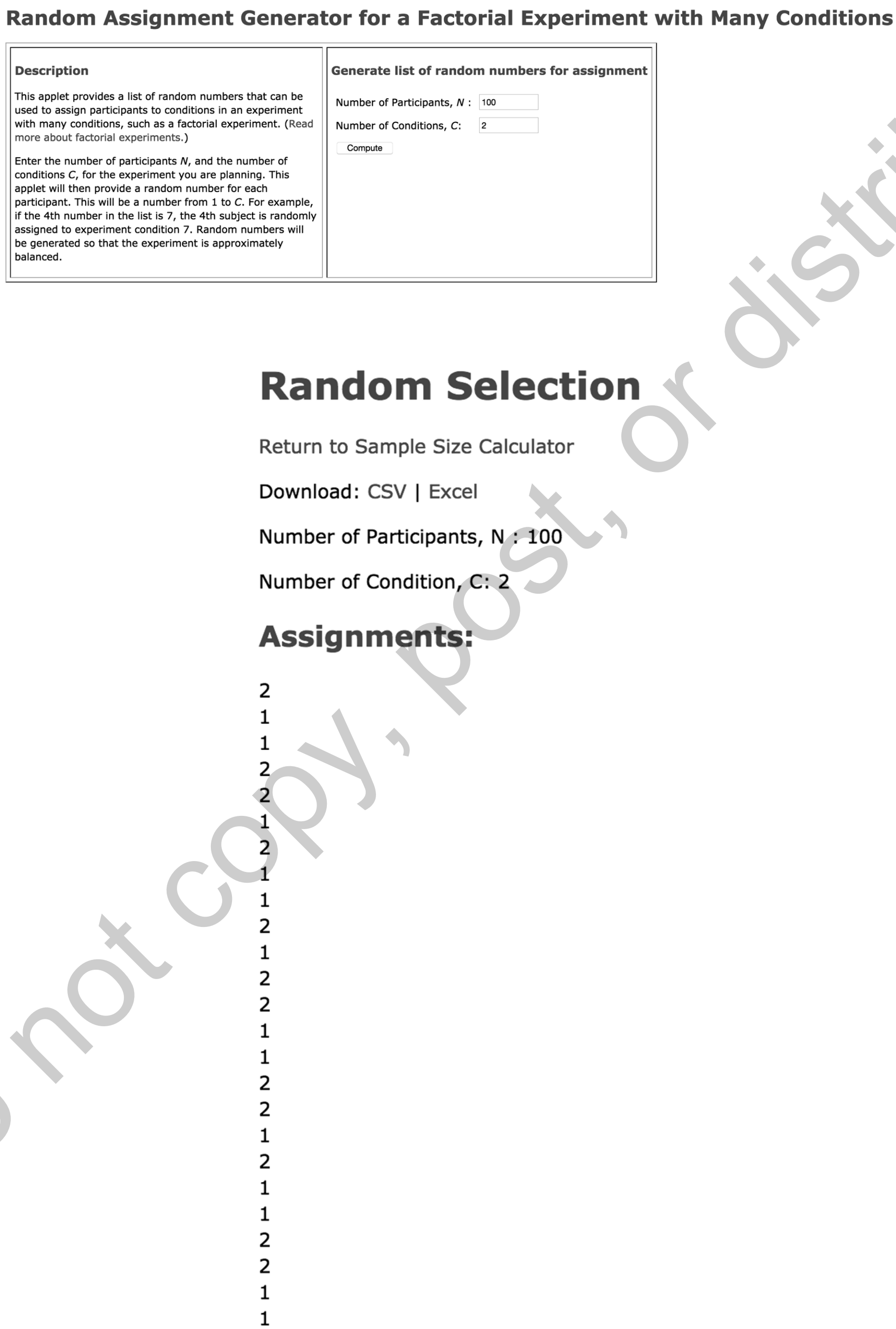
FIGURE 7.1 ● GRAPHPAD



Subject #	Group Assigned
1	A
2	A
3	B
4	B
5	B
6	A
7	B
8	B
9	A
10	A
11	B
12	A
13	A
14	B
15	A
16	A
17	B
18	A
19	B
20	A
21	B
22	B
23	A
24	B
25	A
26	A
27	B
28	A
29	A
30	A

Source: http://www.graphpad.com/quickcalcs/randomize1.cfm

FIGURE 7.2 ● RANDOM ASSIGNMENT GENERATOR



Random Assignment Generator for a Factorial Experiment with Many Conditions

Description

This applet provides a list of random numbers that can be used to assign participants to conditions in an experiment with many conditions, such as a factorial experiment. (Read more about factorial experiments.)

Enter the number of participants *N*, and the number of conditions *C*, for the experiment you are planning. This applet will then provide a random number for each participant. This will be a number from 1 to *C*. For example, if the 4th number in the list is 7, the 4th subject is randomly assigned to experiment condition 7. Random numbers will be generated so that the experiment is approximately balanced.

Generate list of random numbers for assignment

Number of Participants, *N* : 100

Number of Conditions, *C*: 2

Compute

Random Selection

Return to Sample Size Calculator

Download: CSV | Excel

Number of Participants, N : 100

Number of Condition, C: 2

Assignments:

2
1
1
2
2
1
2
1
1
2
1
2
2
1
1
2
2
1
2
1
1
2
2
1
1

Source: http://www.methodologymedia.psu.edu/most/rannumgenerator

Online randomizers will automatically assign equal numbers of subjects to each group if the maximum number of subjects is divisible by the number of groups; for example, for three groups with forty subjects in each, set the total number of subjects at 120, not 125.

Survey Experiments

One of the newest techniques revolutionizing the way experimental subjects are randomly assigned is with software designed to administer surveys online, such as Qualtrics and SurveyMonkey, among others. These **survey experiments**, also frequently described as an "experiment embedded in a survey,"[23] use a randomizer function to randomly assign subjects to treatment and control conditions without the researcher having to do anything other than set the randomizer before launching the survey. Computer-Assisted Telephone Interviewing software can also randomly assign subjects to conditions, similar to the survey software, with subjects taking the study over the phone instead of online.

Some researchers reserve the term *survey experiment* for studies that randomly sample from a population and also randomly assign subjects to conditions,[24] while others use the term when random sampling is not used[25] but random assignment is. Erisen[26] and colleagues make a distinction between survey experiments and lab experiments, explaining that lab experiments are conducted in a controlled environment where factors such as room temperature, time of day, or other things that could contaminate results can be controlled, whereas survey experiments can be taken by subjects in their own homes and with external factors out of the control of researchers, such as whether the subject takes a break to answer the door, go to the bathroom, or take a phone call. This is especially important when measuring things like the time a subject takes to complete the study, knowledge questions where subjects could look up the answers online,[27] or group decision studies where subjects may not believe they are interacting with real people over the Internet.[28] In cases where this kind of internal validity needs to be assured, survey software can be used with subjects in a lab, providing researchers with both control and the convenience of random assignment and data collection by computer. In a study of politicians' decision making,[29] the researchers had subjects do the study online using survey software, but in order to maintain control, they performed the study in the presence of a researcher. Other researchers have explored ways to discourage cheating on knowledge questions by looking up answers so that online experiments are less prone to this kind of error.[30]

There has been much debate and research on the subject of whether survey experiments that use random samples are better because they are generalizable, or if convenience samples produce similar results. Many studies show little to no differences.[31] That issue was covered extensively in this book in the section on external validity in chapter 5. The point in this chapter is that whether it is called a survey experiment or lab experiment, the defining feature of all true experiments is that subjects must be randomly assigned to treatment and control groups.

In-Person Randomizing

Having subjects participate in an experiment via computer makes random assignment easier on many levels. But sometimes an experiment is best conducted with paper and pencil. In-person studies require different random assignment strategies. For example, political scientists might conduct an experiment using voters as they exit polling stations to ensure that subjects actually voted. If a message in a pamphlet or newspaper is the message being tested, then using a printed version is more realistic than one shown online. Most studies on moral judgment are conducted in person because of the complexity of the topic.[32] Moral judgment experiments with journalists may be conducted in the newsrooms with the researcher providing lunch and having subjects take the study while they eat; this generates more participation from busy working professionals.[33] Other experiments may be conducted at professional association conferences, giving the study with paper and pencil at tables in the lobby.[34] For some studies, an old-fashioned paper-and-pencil test conducted in a classroom or library might provide faster data collection.[35] In these cases, random assignment needs to be conducted by online randomizers or old-fashioned methods described earlier.

REPORTING RANDOM ASSIGNMENT

Regardless of how random assignment is achieved, it should always be reported in the research paper as having been done, and preferably the procedure used. No great amount of detail on the specifics is usually necessary, but the report should at least say that subjects were randomly assigned to conditions lest readers think otherwise. Here are two actual published examples:

> "Participants were randomly assigned to a condition."[36]

> "Participants were then randomly given a booklet, which contained instructions for the research, a test pamphlet, and a response questionnaire. The test pamphlet was one of the four versions of a health pamphlet about HPV and genital warts. Thus participants were randomly assigned to one of the experimental conditions."[37]

The best practice is to provide a description of the randomization mechanism—for example, whether it was done with a deck of cards or die, or the randomizer function in a particular piece of software—and whether it was pregenerated or produced on site.[38]

BALANCED AND UNBALANCED DESIGNS

Random assignment can raise concerns about unequal numbers of subjects in groups, as it did for me in my first experiment. This is called an **unbalanced design**. Very unequal sample sizes can affect group equivalence. It is not crucial to have exactly the same number of subjects in each group; as long as the numbers in each group are close, it will be approximately balanced.[39] However, balanced designs, where the exact same numbers of subjects are in each group, give a study more statistical power (the subject of chapter 8).

There may be very practical and unavoidable reasons for an unbalanced design, such as some subjects dropping out or being purged. For example, in moral judgment studies, there is a built-in check for subjects who are trying to fake a better score.[40] When faking is detected, these subjects are eliminated from the data set, usually resulting in unequal numbers of subjects in groups. Sometimes, researchers lose subjects who did not complete enough questions for their data to be useful, or when subjects systematically drop out of a study, which is called **attrition**. Concerns about unequal numbers of subjects in groups should never be a reason to deviate from randomization, as failing to properly randomize is a far greater threat than unbalanced groups. As was pointed out in the discussion of control groups in chapter 6, having up to a third fewer subjects in the control group of an incomplete factorial is acceptable.[41] When treatments are expensive or difficult to run, having fewer subjects in the groups that are of less interest is also acceptable. In addition, when a treatment is desirable, people may be reluctant to participate if they think they may be denied it by being assigned to the control group. Having more subjects randomly assigned to the treatment group can help overcome these objections.[42]

It is also important to report attrition rates, as was done in a business experiment of a mentoring program. In that study, 52% of subjects dropped out of the study; the researcher compared the dropouts to those who remained in the study and found no statistical differences on the observable characteristics.[43] Rules of thumb for attrition rates say between 5% and 20% may be a source of bias.[44]

Many good statistics books explain ways of analyzing data from unbalanced designs. Statistical techniques such as Levene's test[45] can be used to determine if the unequal number of subjects results in unequal variance. When the Levene's test is significant, indicating the variance is not homogeneous, the researcher then uses more stringent tests of differences that do not assume equal variances. This is a subject for a statistics class or text. Suffice it to say that there are better ways to deal with unbalanced groups than by going off course with the randomization mechanism.

CHECKING THAT RANDOM ASSIGNMENT WAS EFFECTIVE

Even though random assignment is the best method researchers have for getting equivalent groups, and randomization failure is rare,[46] some are skeptical. That is when a random assignment or balance check, which is a comparison of the groups' equivalence, is in order. It is not necessary for *all* variables to be equally distributed among groups; those that are highly related to the outcome or dependent variables are of most concern. A rule of thumb for when something is "highly related" is that if a variable correlates with the dependent variable at .45 or greater, that variable must be equivalently distributed among groups.[47] It is a common misunderstanding that random assignment must result in equivalence on every variable known to mankind; that is not the case.[ii]

Aggregate Level Random Assignment

Random assignment checks are more important when assignment is done at aggregate levels rather than with individuals; for example, businesses, families, polling precincts, or classrooms can be randomly assigned as units.[48] Aggregate level assignment[iii] is usually done when it is difficult or impossible to randomly assign individuals, but this is less preferable than individual random assignment, as people within a group or organization may differ systematically. For example, a start-up business may have younger workers than an established firm. An intact classroom may have group dynamics that create different motivation levels. Polling precincts may have voters who vary systematically because of the neighborhood they can afford to live in. It is more important to test equivalence when groups, such as classes or whole schools rather than individuals, are randomly assigned. This is because the sample size is usually lower when using aggregate groups than when using individuals.[49] For example, one study[50] assigned forty-one schools to three conditions, giving each group thirteen to fourteen schools. This is well below the suggested size of twenty.[51]

Reporting Random Assignment Results

Not all journal articles will report the results of a random assignment check, but the practice is growing more common, although not without controversy.[iv] The field of political science has developed guidelines that require reporting whether random assignment was employed, as well as the unit that was randomized, and tables or text showing baseline means and standard

[ii]For a detailed discussion of this, see Mutz and Pemantle.

[iii]It is also important that the unit of *assignment* be the same as the unit of *analysis* in statistical tests. For example, if schools or classrooms are assigned as aggregate units to receive a treatment, then statistical analysis should be based on the aggregate level, not the individual level. In a hypothetical study of 2,500 students at thirty schools, fifteen of which are assigned to treatment and fifteen to control conditions, the total sample size is thirty, not 2,500. Otherwise, the precision of results will be overstated due to the overinflated N.

[iv]To get a sense of the controversy, read Gerber et al., (2014), the challenge to it by Mutz and Pemantle (2016), and Gerber et al.'s response (2016): Alan Gerber et al., "Reporting Guidelines for Experimental Research: A Report from the Experimental Research Section Standards Committee," *Journal of Experimental Political Science* 1, no. 1 (2014): 81–98; Alan S. Gerber et al., "Reporting Balance Tables, Response Rates and Manipulation Checks in Experimental Research: A Reply from the Committee That Prepared the Reporting Guidelines," ibid. 2, no. 2 (2016): 216–29; Diana C. Mutz and Robin Pemantle, "Standards for Experimental Research: Encouraging a Better Understanding of Experimental Methods," ibid.

deviations for certain variables.[52] To monitor the results of random assignment, researchers can test the equivalence of the groups on important variables with means and standard deviations or statistics designed to detect differences such a *t* tests, chi-square, and analysis of variance (ANOVA). These are then reported in a table. For example, if gender is important to the outcome variable, it will be checked to see that the groups have equivalent numbers of men and women; this is one time where finding no significant difference is a good thing. Only basic information is necessary for most randomization reports. For example, here is how one study reported it in an experiment testing the vividness effect on health messages:

> Two-way ANOVA crossing the two manipulated variables ("message vividness" and "argument strength") were performed on participants' gender, age, sex behavior, and so forth. Results showed that the participants in the four experimental conditions were not significantly different from each other ($p > .05$). Therefore, randomization appears to be effective.[53]

Some authors go further. For example, one study[54] included a table showing the descriptive statistics for various independent variables by group but no significance tests (see figure 7.3). Here is the narrative that was included in that study, and also the table:

> The means of the independent variables in each of the experimental conditions are reported in table 1. As shown in table 1, there were no substantial differences between the means and standard deviations of all the independent variables. The gender proportions in both conditions were also equivalent. Since the independent variables were asked before the manipulation, this shows that the random assignment indeed resulted in equal groups.[55]

Ho and McLeod included this table illustrating that random assignment resulted in equivalent groups on various variables.

Here is another example from a study on tutoring that did include the results of significance tests of random assignment and also included a table:

> In order to test whether the students were distributed randomly in terms of these background measures, group means were calculated and t-tests were run to test for significant differences between the treatment and control groups. Tests of significance indicated that, for all background measures but one, there were no statistically significant differences between the treatment and control groups: the randomisation had worked to create pre-treatment equivalence.[56]

It is important to report randomization checks after a study has been completed if there has been considerable attrition, or subjects dropping out. This is to ensure that the dropouts did

FIGURE 7.3 ● TABLE OF RANDOM ASSIGNMENT

	FTF (*n* = 192)		CMC (*n* = 160)	
	M	*SD*	*M*	*SD*
Gender	Females (71.9%)	—	Females (69.1%)	—
Print news use	4.73	2.07	4.49	2.11
Television news use	5.02	1.96	4.98	2.04
Fear of isolation	2.98	1.10	3.26	1.10
Communication apprehension	3.00	1.23	3.03	1.26
Current opinion congruency	.09	.24	.09	.25
Future opinion congruency	.32	.50	.30	.49

From: Ho, Shirley S., and Douglas M. McLeod. 2008. "Social-Psychological Influences on Opinion Expression in Face-to-Face and Computer-Mediated Communication." *Communication Research* 35 (2) (April): 190–207.

not differ significantly from the subjects who stayed in the study. Sometimes, the treatment itself is the reason. It may be too long, boring, or difficult, so subjects quit. Experiments that are conducted over a long period of time are especially prone to this problem.

When Random Assignment Fails

Random assignment is the best available method for achieving equivalence among experimental groups, but it does not guarantee that groups will be perfectly matched on every individual difference variable.[57] It minimizes, rather than prevents, confounding.[58] Differences still may occur due to chance. It is not necessary to have equivalent groups on every possible variable; it is most important on variables that are correlated with the outcome variable.[59] For example, in moral judgment studies, age and education are highly correlated with moral judgment, but gender is not.[60] So it is more important to have groups be equivalent on age and education, and not worry so much about having equivalent numbers of men and women. When groups are not equivalent on important characteristics, internal validity decreases and researchers run the risk of making erroneous inferences.[61] However, failure to achieve equivalence with a proper randomization mechanism is rare.[62] Furthermore, the alpha level of significance testing already takes into account the fact that some variables will be spread unevenly across groups due to chance.[63] There is no way to "fix" a true failure of random assignment other than to start from scratch and redo the randomization properly.[64]

FIGURE 7.4 ● RANDOM ASSIGNMENT REPORTING

	FTF (n = 192)		CMC (n = 160)	
	M	*SD*	*M*	*SD*
Gender	Females (71.9%)	—	Females (69.1%)	—
Print news use	4.73	2.07	4.49	2.11
Television news use	5.02	1.96	4.98	2.04
Fear of isolation	2.98	1.10	3.26	1.10
Communication apprehension	3.00	1.23	3.03	1.26
Current opinion congruency	.09	.24	.09	.25
Future opinion congruency	.32	.50	.30	.49

From: Ritter, Gary W., and Rebecca A. Maynard. 2008. "Using the Right Design to Get the 'Wrong' Answer? Results of a Random Assignment Evaluation of a Volunteer Tutoring Programme." *Journal of Children's Services* 3 (2): 4–16.

Here are some tips for achieving equivalence:

- If it is possible to include all subjects in all the groups without carry-over effects—that is, using a within-subjects design—then equivalence is a nonissue because the exact same people are in each group.
- Groups are considered nonequivalent if twice as many subjects in one group have some nuisance variable compared to the other group[65]—for example, if there are twice as many men as women in one group where gender is related to the outcome; ten men and five women in a group is considered nonequivalent.
- Start by ensuring there are enough subjects in each group. A power analysis, explained in chapter 8, will do this. The more subjects, the greater the chance that equivalence will be achieved.[66] In studies that employ multiple runs, more subjects can be recruited and randomly assigned, and the experiment conducted again to reach equivalence.[67]
- A good rule of thumb for achieving equivalence is to have at least twenty subjects in each group.[68] But even then, groups with fewer than twenty subjects actually are protected from erroneous conclusions of nonequivalence because statistics tests have a harder time detecting spurious differences with small numbers.[69]
- The random assignment process can be redone until equivalence is achieved.[70] Obviously, this only works if random assignment can be checked before the treatments are given.

- If nonequivalence can only be detected after the study is conducted, one widespread strategy is to use the nonequivalent variable as a covariate in the statistical analysis using analysis of covariance.[71] This will help reduce the influence of that variable before testing differences between the groups. For example, if there are twice as many men in the treatment group as the control group, and gender is expected to affect the outcome, using gender as a covariate will help level the playing field. That is, it helps remove unexplained variability due to the effects of gender before testing the effects of the treatment, which leads to more precise estimates.[72] This approach should be used conservatively, however, as covariates should only be used if planned in advance (a topic of the next chapter), as this will not "control" for the lack of true random assignment. And the growing tendency to include numerous covariates defeats the purpose of a well-designed, controlled experiment.[73]

One thing researchers should *not* do is purposively recruit more subjects with the needed characteristic and assign them just to the group low on that variable.[74] That is not random. Nor should a researcher try to rebalance the groups by moving subjects around to even out the groups before giving the treatment.[75] That also is not random.

Finally, it is somewhat reassuring to know that if groups are not equivalent on some variables, the differences will represent random error, not systematic error, and are unlikely to produce incorrect inferences.[76] Additionally, replication helps correct for any erroneous conclusions from a study threatened by nonequivalence; for more on that, refer to chapter 5. Over multiple studies, the truth tends to prevail.[77]

In fields where equivalence can be prone to problems, such as in education, social work, criminology, and program evaluations, a significant amount of time may need to be devoted to achieving equivalent groups. For example, in a study of a drug program in schools,[78] after random assignment was completed and checked for equivalence, two schools dropped out, making the groups nonequivalent. The researchers had to draw new schools, randomly assign them, and recheck equivalence.

BLOCKING, MATCHING, AND OTHER STRATEGIES

One way to reduce the chances of unequal groups is with a **matched pairs** strategy and **blocking**. This involves matching subjects on important variables and then assigning them to treatment and control groups as a pair or block.[79] This is used to help ensure that an extraneous or nuisance[80] variable related to the outcome variable does not confound the results. Groups are created based on subjects with the same level of the blocking variable. For example, if gender is the blocking variable, subjects would be randomly paired by gender—a man with a woman—and then each pair randomly assigned to the treatment or control group.[81]

In a business study on the effectiveness of a mentoring program, the researcher was not allowed to randomly assign subjects to condition by the employers, who wanted to select employees with the highest potential for promotion to the program.[82] Instead, the study used a matched pairs design, selecting control group employees who were similar to treatment group employees based on five characteristics such as similarity of salary, performance rating, tenure in the organization, working in the same office, and not previously participating in a mentoring program. The study reports statistical tests showing no differences between treatment and control group subjects on these variables, while also noting "the treatment and matched control groups may have varied on unobserved characteristics."[83]

Experimenters must anticipate and be able to measure the variable before conducting the study, so matching is no help for confounding variables that are not known; simple random assignment is still preferable for this reason.

Blocking also can be done with more than two levels of variables. For example, if age is the variable one wanted to ensure is equivalent across groups, then blocks of different age groups could be created; for example, four blocks for the age groups eighteen to thirty-four years old, thirty-five to fifty-four years old, fifty-five to sixty-four years old, and sixty-five years and older. After these blocks are created, the subjects in each are randomly assigned to treatment and control groups so that an equal number of subjects from each age block are in each group. Now, age cannot be the cause of any differences between the experimental groups.

Blocking is not preferable to straightforward random assignment but is useful when very few potential subjects are available or small sample sizes are likely—for example, when groups such as schools are the units. Matching strategies are frequently used in education studies, where schools are matched on important characteristics and then randomly assigned one to each condition.[84] Another example comes from a study on the effects of having grown up with friends of a different ethnicity on stereotyping.[85] The researcher anticipated it would be hard to find subjects who had grown up in integrated neighborhoods, so he pretested experimental volunteers, measuring their level of personal contact with minorities, and then matched high-contact subjects with low-contact subjects before randomly assigning them to the treatment or control group.

Blocking and matching strategies also can be useful when random assignment is not likely to be implemented correctly—for example, in program evaluations where the researcher is not in control of assignment.[86]

Blocking vs. Simple Random Assignment

Matching strategies are *not* preferable to simple randomization,[87] although they can be misleading in their intuitive appeal.[88] For one thing, statistical tests lose power when blocking factors do not have much influence on the outcome variable.[89] Blocking is more common in the life sciences when studying plants and animals.[90] Another drawback is that blocking requires a two-step process, first measuring subjects on the blocking factor, then randomly assigning them to groups, administering the treatment, and measuring the outcome. Social science experts agree that simple random assignment is preferable to other methods for achieving comparable groups.[91] Even the harshest critics of random assignment do not argue for an alternative.[92] Campbell and Stanley are particularly critical of the "widespread and mistaken preference . . . for equation through matching,"[93] saying, "matching is no real help when used to overcome initial group differences." When confounding variables are unknown, and so uncontrollable (sometimes called *lurking variables*), random assignment is the best strategy, as it automatically balances these.[94]

Stratified Random Assignment

One technique that makes it easier to assess equivalence on many variables is **stratified random assignment**, where numerous variables are combined into a single variable, similar to a factor created by factor analysis.[95] This does not block or match up subjects or units using any particular variables, but rather combines numerous related variables into one overarching factor that can be measured for equivalence after subjects are randomly assigned. In other words, it allows researchers to measure equivalence on one global factor instead of worrying about numerous discrete variables. For example, one study[96] used seven key variables such as the type of community a school was in, number of grades in the school, percentage of White students, enrollment per grade, and rural or urban setting. It then used a statistical technique to arrive at one composite stratifying variable. This was done to find a combination of variables that were closely related within these schools, which the authors called "rurality," explaining that rural schools often were similar on these characteristics. After random assignment, equivalence was checked by the single composite factor rather than on seven individual variables. The actual process was more complicated than described here.[97] The authors found equivalence using ANOVA tests of differences and reported it in a table (see figure 7.6). This stratifying procedure has the advantage of having unknown or unmeasured variables be randomly distributed across groups, whereas blocking and matching strategies do not.[98] As with achieving equivalence, this strategy is reserved for variables that are likely to be highly correlated with the outcome variable, not for every variable.[99] This technique, like blocking, is more common in some disciplines than others, so it is important to know the standard in your field.

RANDOM ASSIGNMENT OF OTHER THINGS

So far, this chapter has focused on randomly assigning individual subjects to conditions. But random assignment applies to more than simply how subjects are assigned to groups. In fact, experts advise randomly assigning as many of a study's procedures as possible.[101]

FIGURE 7.5 ● RANDOM ASSIGNMENT CHECK

The text explaining the random assignment equivalence check said, "First, we verified that the assignment procedure worked to provide group equivalence on the school-level variables relating to CIS score. We conducted a simple ANOVA (SAS Proc GLM), with Program (Rural, Classic, Control) listed as a class variable. Table 3 presents the results for the overall F test. As expected, program group membership was unrelated to the CIS score, the two factors making up the CIS score, and the seven items making up the factors."[100]

	Tutored students	Control students	Total sample
Average reading grade, 1997-98 year-end (A = 4.0)	1.65	1.61	1.63
Average math grade, 1997-98 year-end (A = 4.0)	1.79	1.85	1.82
Average SAT-9 reading open-ended national percentile score, 1998	23.1	23.2	23.2
Average SAT-9 math open-ended national percentile score, 1998	16.7	18.6	17.6
Full year attendance rate, 1997-98	91.8	90.6	91.2
% of students not promoted, 1997-98	9.2	9.0	9.1
% of students African American	95.9	96.8	96.4
% in home receiving welfare assistance (TANF)	63.3	60.9	62.1
% with guardian with a high school degree or more	68.6	74.5	71.5
% with guardian currently working for pay	57.2	51.9	54.6
% with guardian reporting health problem that limits activity	19.5	23.2	21.3
% in home with both mother and father	37.4	31.1	34.4
% reporting that someone helps with homework	73.2	78.4	75.7
% reporting that someone at home reads with them	67.0	69.5	68.2
Average number of children in household	3.39	3.02	3.20
Total study sample	**196**	**189**	**385**

1. The welfare assistance data were derived from the School District of Philadelphia student information system.
2. The background data related to student guardians and the numbers of children per household were derived from the baseline survey completed by the guardians who gave parental consent for the student to participate in the tutoring programme (September 1998).
3. The figures on household composition and help with reading and homework were derived from the tutee follow-up survey administered to tutees at programme completion in May 1999.

Shadish, Cook, and Campbell deliberately refer to random assignment of "units" so as not to imply that only people should be randomized.[102] Anything that could introduce systematic bias should be randomly assigned. For example, if more than one experimenter

FIGURE 7.6 ● TESTING FOR RANDOM ASSIGNMENT TABLE

Variable	*F*(2,38)	*p*
CIS	0.25	.78
Factor 1	1.05	.36
Factor 2	0.08	.93
Rurality	0.45	.64
Npergrade	1.62	.21
Numgrades	0.41	.67
Pctwhite	1.60	.21
Pctlunch	0.10	.90
Scores	0.22	.81
Rdrugs	0.80	.46

Note: Post hoc tests with the Duncan test showed no significant differences ($p<.10$) for any individual comparisons.

will administer the study, the experimenters should be randomly assigned to the sessions and conditions he or she will supervise.[103] An experimenter who observes and measures subjects might become better after practice or, conversely, grow tired and get worse at measuring. It is important that the experimenter not be assigned to observe all subjects in the control group first and the treatment group last (or vice versa) in order to avoid introducing systematic differences between the groups. Instead, experimenters should be randomly assigned to each group as well as to session.

Typically, when there is more than one stimulus—advertisements, news stories, and health messages, for example—these should be presented to subjects in a random order. In a study of politicians' decision making, the researchers randomized many things.[104] In addition to randomly assigning subjects to condition, they randomized the three types of tasks and also the thirteen decisions within each type of task that subjects had to make. The study also gave subjects an incentive in the form of a donation to a charity, and the way that was offered was also randomized; here is the explanation:

> To make each decision that includes a monetary payoff relevant, but simultaneously ensure that tasks do not influence each other, we randomly selected one task that determined how much money was donated to the charity on behalf of the participant. Specifically, we randomly selected either the lottery-choice or the lottery-valuation part of the experiment, and then we randomly selected one task from this part. This avoided that participants' choices were influenced by so-called portfolio effects (e.g., some safe and some risky choices for a balanced portfolio) or by previous earnings.[105]

This study used a within-subjects design, but because the scenarios in the two conditions (gain frames, loss frames) were similar, the researchers did not want subjects to read each scenario in both conditions, so they randomly assigned each participant to read half the scenarios in each condition. They explain it like this:

> For each scenario, it was randomly determined whether a participant was presented with the loss or the gain frame, so it was possible that a participant would get the gain frame for one scenario and a loss frame for another scenario. We also randomly determined the order of the scenarios.[106]

Clearly, these researchers followed Bausell's advice: "When in doubt—randomize."[107]

Counterbalancing

The reasons for randomly assigning or rotating the order of something, or **counterbalancing**, is to avoid carry-over effects. These were described in chapter 6 as effects due to learning, practice, fatigue, or the subjects changing. Carry-over effects happen when receiving one treatment affects a subject's response to the next treatment. One specific kind of carry-over effect arises from the order in which things are presented, known as **order effects**. Order effects have been well documented under the study of **primacy** and **recency**, the ideas that we remember best what we are exposed to first and last. This is a special concern for within-subjects designs where every subject gets all the different treatments, as they are especially prone to fatigue, practice effects, carry-over, and order effects.[108] Counterbalancing is helpful because often the carry-over effect in one direction will cancel out the effect in the other direction. For example, some subjects may do better on the last treatment (recency effect) and others on the first (primacy effect). When the data are aggregated, these two effects cancel each other out. The same applies to everything that is randomly assigned; for example, if more than one experimenter will supervise multiple runs of a study, the different experimenters should not only be randomly assigned to the treatment or control conditions, they should also be rotated, or counterbalanced, among sessions.[109] The goal of counterbalancing is to make order effects equivalent across the conditions. And, as with all things equivalent, failing to counterbalance orders decreases internal validity.

Latin Square

Counterbalancing can be accomplished by simply randomly assigning, but there is one specific type of counterbalancing strategy used in experimental research called the **Latin square** that equalizes the number of positions under which each stimulus occurs. It ensures that each experimental message or stimulus occurs in the first position one time, in the last position one time, and in each in-between position one time. Also, each condition or stimulus follows the other exactly once. This is more efficient than simple random assignment.[110] Here is an illustration of how it works: A hypothetical experiment uses three different advertisements as stimuli. If the researcher randomly assigns the ads to order, there are six possible order combinations:

A	B	C
A	C	B
B	C	A
C	B	A
C	A	B
B	A	C

But using Latin squares produces three combinations:

A	B	C
B	C	A
C	A	B

Each advertisement is shown first one time, last one time, and in the middle one time. The two key features of a Latin square are that a row or column never contains the same letter twice and that every row and column contains the same letters. This is much more efficient than simple random assignment to order,[111] as it reduces the number of subjects needed to receive each order.[112] Latin squares are especially useful in large factorial designs where it would be quite costly to administer all possible order combinations.[113] This type of counterbalancing is achieved by randomly selecting stimuli to represent A, B, and C in the first row.[114] Next, the stimuli are simply rotated by moving the first-place stimulus to last place in each row and sliding the others over one place. This works for three or more stimuli; obviously with two stimuli, there are only two possible order combinations.

In reporting Latin squares, as in reporting random assignment, it is common to just see it mentioned in passing. For example, in this study of public service announcements (PSAs) in a health study, the authors said, "The order of the presentation of PSAs was counterbalanced according to a diagram-balanced Latin square."[115]

And in another study of decision making in TV newsrooms, the author described the rotation of the three story scenarios this way: "The order of which story participants received was counterbalanced using a Latin squares design."[116]

In studies that use multiple factors, Latin squares help avoid confounding of the factors. For example, in a study of the use of humor in advertising,[117] the authors had subjects listen to radio broadcasts that featured different advertisements. The factors used to make up each ad consisted of the type of product (e.g., cereal, cheese, batteries), brand name, and jokes, which they called one-liners. Here is how they describe their counterbalancing strategy:

> Rotations for the three versions of radio broadcasts were designed such that a particular one-liner was not presented with a particular brand name or product type more than once in the study. . . . Within each combination, product types, brand names, and one-liners were rotated in order or presentation for each of the three audiotapes. To arrange three commentaries in three rotations, a Latin square counterbalancing technique (Keppel, 1991) was used.[118]

The Latin squares design got its name from an ancient puzzle on the different ways Latin letters could be arranged in a square.[119] It was introduced as a rotation experiment,[120] popularized by R. A. Fisher,[121] and has been the preferred method in psychology ever since.[122]

RANDOM ASSIGNMENT RESISTANCE

Most researchers are quickly persuaded of the abilities of randomization to solve a multitude of problems, but that is frequently not the case for nonresearchers whose participation may be needed in a study. Disciplines that perform program evaluations or conduct studies in real-world or field environments may encounter resistance to random assignment. For example, school administrators allowed some students to bypass the random assignment process in one study.[123] Education researchers, for example, find that school personnel may object to offering some students a treatment opportunity and denying it to others.[124] Business researchers have found that executives may refuse to allow their employees to be randomly assigned, instead insisting on handpicking those assigned to each group themselves.[125] In a health study, "a substantial number" of subjects refused to participate in the randomization because they did not want to be involved if they were not assured of receiving the potentially beneficial treatment.[126] Ethical concerns also raise opposition to randomization; for example, in an experiment on treatment for crime victims, some subjects who exhibited self-harming tendencies were moved from the control group to the treatment group, disrupting the randomization.[127] This can lead to researchers abandoning a true experiment in favor of a quasi experiment, discussed in chapter 4. This topic is discussed in more detail in More About box 7.3.

As elegant a procedure as random assignment is in all its forms, it is not perfect. But as Campbell and Stanley say, "It is nonetheless the only way of doing so, and the essential way."[180]

With a firm understanding of how subjects (and other things) should be assigned, the next chapter will deal with ways to sample subjects and how to determine the optimum number of them.

MORE ABOUT . . . BOX 7.3

Resistance to Random Assignment

When it comes to random assignment, compliance is critical.[128] As with pregnancy, there is no such thing as being "a little bit random"[129]— it is an all-or-nothing proposition. That distinction is often lost on nonresearchers; for example, in one study, when told that random assignment had been compromised, staff implementing it acted surprised and responded that they believed the assignments were "in fact, more-or-less random."[130] One employee directly involved with the process believed the instructions were merely "recommendations."[131]

Another example of random assignment gone wrong includes a murder by a subject who dropped out of a program, which led the prosecutor to refuse to deny treatment to anyone in need.[132]

Not all threats to random assignment are this dramatic. Any actions that compromise randomization can undermine the internal validity of an experiment by leading to nonrandom differences between subjects in treatment and control conditions. In the following table, the first column lists some common objections to random assignment from practitioners in the field tasked with assigning subjects to conditions for researchers. The column on the right contains advice for overcoming these objections. Researchers who have reported their results with these techniques have seen compliance with randomization go from as low as 19% to 94%.[133]

The advice here is drawn from education, counseling, criminal justice, and business fields but applies to attempts to assess the effectiveness of an intervention in a field setting in any discipline.

The best approach is for researchers to insist on conducting random assignment themselves and to carry out the procedure at the researcher's site, not the study site.[134] This should always be done in conjunction with communication with the site staff, allowing them to voice concerns, have questions answered, and participate in the design process.[135] Monitoring the assignment process is also essential. Subjects have been known to try to sneak into groups to which they are not assigned.[136] It is also important to observe the extent to which the intervention is actually being implemented; for example, if teachers are supposed to use technology, check to see how much they are doing so.[137]

OBJECTION	RESPONSE
FAIRNESS Giving a perceived benefit, even if unproven, to some and not all is intrinsically unfair.[138] School personnel, in particular, are not inclined to offer interventions to some and deny others the same opportunity. [139]	Explain that when resources are limited and there are more people who need services than slots available, random assignment is one of the fairest ways to distribute services.[140] Point out that random assignment protects the organization from accusations of favoritism.[141] Explain the random assignment process as a lottery, which gives everyone an equal chance, which is fair.[142] Alter the control group so they receive a lower dose of the treatment instead of no treatment.[143] Agree to give those in the control group priority to participate in the treatment group during another run of the experiment.[144]

(Continued)

(Continued)

OBJECTION	RESPONSE
NEED Fears that the most in need of the treatment would not be chosen by random assignment.[145]	Offer to categorize some of the most in need as "wildcards" who can bypass the random assignment process and are put in the treatment group; then exclude these subjects from the analysis.[146] Divert the most needy into an alternative intervention that is not part of the study.[147] Explain that the treatment has not yet been shown to work—that is what the study is for. If it does not provide benefit, then the needy in the control group will not have lost anything. [148] Explain that if the treatment is shown to work in this randomized experiment, everyone can be offered it later.[149]
EVALUATION FEARS If the experiment shows the program has no impact, it could be eliminated.[150]	Be sensitive to this issue and collaborate on the outcome variables. Add in qualitative data that can provide insights beyond the quantitative. Also include measures that may show more sensitivity to the program.[151]
CONFLICTS WITH PRACTICE Randomly assigning people to get special training or be enrolled in certain classes conflicts with normal procedures.[152] It may be difficult to assign certain interventions to only parts of existing units. For example, in schools where classes are taught by teams of teachers, assigning an intervention to only one set of teachers can lead to group planning issues.[153] In one study, special education students needed to be in the same class, resulting in special education students being overenrolled in one condition.[154] Staff is uncomfortable with an outsider telling them what to do.[155] It might send unintended signals to high-performing employees if they were not chosen for the treatment.[156] Staff may object to the extra work of implementing random assignment, experience scheduling conflicts, or key people may leave due to illness or turnover.[157]	Get to know staff, the environment, and their needs.[158] Suggest procedures that afford a minimum of disruption.[159] Allow practitioners to exempt up to 10% of subjects from random assignment for practical reasons. Track them and exclude from the analysis.[160]

OBJECTION	RESPONSE
ACCIDENTAL Subjects may be accidentally assigned to the wrong condition due to misunderstanding, lack of time to make assignments properly, or staff turnover.[161] In one study of schools, staff did not understand the lists of students already included random assignment to group, and they devised their own method to randomly assign from the lists.[162]	Improve communication.[163] Assign a randomization liaison from the research team to the organization.[164] Schedule multiple meetings and information sessions to explain the research design.[165] Include everyone. For example, in one study, security guards were in charge of checking subjects into the program, so they needed to know why it was important that subjects go to the group they were assigned.[166] Include staff in initial planning sessions for random assignment. [167] Look for staff and organizations that have prior experience with random assignment and research studies. In one study, a staff member involved with random assignment said it took a year and a half before she fully understood the assignment process.[168] Use color-coded forms to make it easier to quickly see to which condition a subject should be assigned. A study of domestic violence treatment trained police officers in random assignment using color-coded report forms.[169]
SUBJECTS OBJECT Subjects request being put in different groups, refuse to participate in the randomization, withdraw in the middle of treatment, or complain to administrators, who reassign them.[170] Subjects drop out for a variety of reasons including health, lack of childcare, and transportation issues.[171]	Make the control group more attractive by giving subjects some sort of "treatment as usual" such as normal classroom instruction or other interesting activities that will not confound the outcome.[172] One set of researchers considered an open-ended discussion group as an enhancement to the basic treatment.[173] Establish a procedure so that those requesting to be moved must meet with the researchers, who explain the reasons behind the assignment.[174] Offer the intervention to the subject at a later time.[175] Put control group subjects who wish to receive the intervention on a waiting list for future programs.[176] In one study, a principal insisted that siblings of those chosen for the treatment group also had to be included. Exclude these subjects from statistical analysis.[177] Have staff refer subjects' requests to researchers, who are in a better position to turn them down and explain why[178] Track subjects who leave and learn why.[179]

STUDY SPOTLIGHT 7.4

Taking advantage of random assignment in a natural setting

From: Abrams, David S., and Albert H. Yoon. 2007. "The Luck of the Draw: Using Random Case Assignment to Investigate Attorney Ability." ***University of Chicago Law Review*** **74 (4) (Fall): 1145–1177.**

This study took advantage of naturally occurring random assignment when the researchers discovered that a county in Nevada was assigning incoming felony cases randomly to attorneys in the pool, which allowed for a natural experiment free from selection bias.

Clark County, which includes Las Vegas, began assigning attorneys to cases after a defendant's death sentence was overturned because he was assigned to an inexperienced public defender. Under the previous nonrandom assignment method, the better attorneys might be assigned the more difficult cases, thus confounding attorney ability with case difficulty. This opportunity allowed the researchers to examine the performance of attorneys that cannot be explained by case characteristics. Conventional wisdom says that lawyers who attend better law schools may get clients lower sentences, for example. The study discovered that Hispanic attorneys and those with more experience achieve better outcomes for clients than others, but gender and law school attended made no difference.

For the purposes of this book, the study is important because it illustrates in depth the value of random assignment.

The researchers began by checking to see that random assignment was indeed being implemented and not thwarted. This helped rule out alternative explanations for trial outcomes, such as case difficulty. Cases were assigned to attorneys without the judges, prosecutor, or team chief knowing any characteristics of the defendant or even the alleged offense, helping to ensure against any subconscious efforts to assign cases purposively. The researchers used nonparametric tests (chi-square) to see if cases were indeed being assigned randomly using the defendants' age, gender, and race. They explain that these three variables are highly correlated with other defendant characteristics on unobserved variables. They say, "Crucially, we assume that this provides evidence that unobservables are also randomly assigned (due to correlation with observables)."[181]

TESTING FOR RANDOM ASSIGNMENT

Case characteristic	*p*-value	Observations
Defendant sex	0.851	10,129
Defendant age	0.253	9,803
Defendant race	0.098	7,145

Note: Each row reports results from a separate simulation to test for the equality of public defender fixed effects. Defendant sex is a dummy variable for whether the defendant is male. Defendant race is 0 for black defendants and 1 for white defendants.

Nonsignificant results showed the cases were indeed randomly assigned (see figure 7.6). This then allowed the researchers to test their hypotheses concerning the differences among attorneys' abilities and other variables that predict better trial outcomes for defendants.

These researchers were creative in spotting an opportunity for a natural experiment. As with most journal articles, this one does not mention the researchers gaining approval from an Institutional Review Board (IRB) before collecting data. The data used may have been exempt from informed consent because it was a matter of public record, or the researchers may have given informed consent after the fact.

New researchers sometimes assume that because data are already collected, this constitutes "secondary data" and one does not need approval from an IRB. True secondary data occurs when subjects have received informed consent—for example, in existing data sets such as the American National Election Studies or Pew polls. The organizations collecting these data sets and making them available to researchers have obtained permission from a human subjects committee of an IRB and have provided subjects with informed consent. This is not the case with all existing data. For example, students who fill out evaluations about satisfaction with their classes have not been given the information contained in informed consent documents. In any research that involves information collected from human subjects, researchers should contact their IRB to determine if they need to obtain IRB approval and subjects' informed consent, even if data have already been collected. Chapter 11 will discuss informed consent and IRB approval in more detail. Researchers should be attuned to opportunities for naturally occurring randomization but should also be careful to secure permission from an IRB and follow protocols for obtaining informed consent from the people whose data will be used.

Common Mistakes

- Not randomly assigning subjects to groups, or not doing random assignment appropriately
- Failing to randomly assign other elements of a study, such as the experimenters, to sessions
- Not counterbalancing stimuli

Test Your Knowledge

1. When a participant has an equal chance of being in the treatment or control group in an experiment, it is called _____.
   a. Random sampling
   b. Random assignment

c. Random error
d. Selection bias

2. The main reason for using random assignment in an experiment is to ensure which of the following?
   a. A sample representative of the population
   b. That neither subject nor experimenter knows which group someone is in
   c. That groups are as equivalent as possible on known and unknown variables
   d. That the dependent variable does not differ across conditions

3. Which of the following is NOT a way to randomly assign subjects to groups?
   a. Drawing names out of a hat
   b. Flipping a coin
   c. Using a random number generator
   d. Rotating subjects so that groups come out even (e.g., 1, 2, 1, 2 . . .)

4. Groups need to be equivalent on all variables that can be measured.
   a. True
   b. False

5. Which is the preferred way to ensure that systematic variation does not confound a study?
   a. Pretesting
   b. Blocking or matching
   c. Simple random assignment
   d. Balancing groups by moving subjects around after random assignment

6. What besides subjects should be randomly assigned?
   a. Nothing, only the subjects
   b. Experimenters
   c. All of the study's procedures
   d. All but A

7. Latin square is a technique for _____.
   a. Creating equivalent groups
   b. Controlling for extraneous individual variables
   c. Minimizing order effects
   d. Randomly assigning subjects to groups

8. Complete the following to make a Latin square:

A	B	C	D

9. If men and women are paired and then assigned to the treatment or control group as a pair, this is called _____.
   a. Blocking or matching
   b. Random assignment
   c. Stratified random assignment
   d. Latin square

10. One drawback to random assignment is that experimenters must anticipate and be able to measure confounding variables before conducting the study, so it is no help for confounding variables that are not known.
   a. True
   b. False

Answers

1. b
2. c
3. d
4. b
5. c
6. d
7. c
8.

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

9. a
10. b

Application Exercises

1. Using the experimental study you began developing in chapter 1 and continued by creating a design table and control group for in chapter 5, decide how you will randomly assign subjects to groups. Write one page on which strategy you will use and why—random number generator, drawing out of a hat, etc. Do a “test run” of this. Assume forty subjects in each of the groups in your study. Use your choice of randomizer to assign subjects. Analyze the results; were they balanced or unbalanced? Repeat the

process with a different type of randomizer to see how it turns out (i.e., if you used an online randomizer first, repeat by drawing out of a hat).

2. Write one page about all the elements of your study that could be randomly assigned. Besides the subjects, what other procedures might be randomly assigned? Why? Assume you will use at least three stimuli in your study (e.g., three different treatments, interventions, teaching techniques, ads, PSAs, stories, messages, etc.). Design a plan for counterbalancing them.

3. Write one page about the possible confounding variables that random assignment will need to help equalize across groups. Read literature about your outcome variable to see what others have found to be highly correlated with your dependent variable. Use your imagination and common sense to identify as many as you can.

Suggested Readings

Chapter 5 in: R. Barker Bausell. 1994. *Conducting Meaningful Experiments: 40 Steps to Becoming a Scientist.* Thousand Oaks, CA: Sage.

Chapter 8, "Randomized Experiments: Rationale, Designs, and Conditions Conducive to Doing Them," pp. 246–278, in: Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth Cengage Learning.

Read these two articles, one critical of random assignment and one more supportive, for a comparative perspective:

- Krause, Merton S., and Kenneth I. Howard. 2003. "What Random Assignment Does and Does Not Do." *Journal of Clinical Psychology* 59 (7): 751–766.
- Strube, M. J. 1991. "Small Sample Failure of Random Assignment: A Further Examination." *Journal of Consulting & Clinical Psychology* 59 (2): 346–350.

Notes

1. W. A. McCall, *How to Experiment in Education* (New York: MacMillan, 1923), 41.
2. R. Barker Bausell, *Conducting Meaningful Experiments: 40 Steps to Becoming a Scientist* (Thousand Oaks, CA: Sage, 1994).
3. Richard A. Berk, "Randomized Experiments as the Bronze Standard," *Journal of Experimental Criminology* 1 (2005): 416–433.
4. Graeme D. Ruxton and Nick Colegrave, *Experimental Design for the Life Sciences,* 3rd ed. (Oxford, UK: Oxford University Press, 2011).
5. Bausell, *Conducting Meaningful Experiments*.
6. D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1963).
7. Charles Sanders Peirce and Joseph Jastrow, "On Small Differences of Sensation," *Memoirs of the National Academy of Sciences* 3, no. 75–83 (1885): 8.
8. William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Belmont, CA: Wadsworth Cengage Learning, 2002).

9. J. M. Gueron, "The Politics of Random Assignment: Implementing Studies and Impacting Policy," in *Evidence Matters: Randomized Trials in Education Research*, ed. F. Mosteller and R. Boruch (Washington, DC: Brookings Institution Press, 2002), 26.
10. Diana C. Mutz and Robin Pemantle, "Standards for Experimental Research: Encouraging a Better Understanding of Experimental Methods," *Journal of Experimental Political Science* 2, no. 2 (2016): 192–215.
11. Martin T. Orne, "Demand Characteristics and the Concept of Quasi Controls," in *Artifacts in Behavioral Research: Robert Rosenthal and Ralph L. Rosnow's Classic Books*, ed. Robert Rosenthal and Ralph L. Rosnow (Oxford: Oxford University Press, 2009), 110–137.
12. Cathaleene Macias et al., "Preference in Random Assignment: Implications for the Interpretation of Randomized Trials," *Administration & Policy in Mental Health & Mental Health Services Research* 36, no. 5 (2009): 331–342.
13. T. D. Cook and D. T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (Chicago: Rand-McNally, 1979).
14. Colin Ong-Dean, Carolyn Huie Hofstetter, and Betsy R. Strick, "Challenges and Dilemmas in Implementing Random Assignment in Educational Research," *American Journal of Evaluation* 32, no. 1 (2011): 40.
15. Macias et al., "Preference in Random Assignment," 331–342.
16. Ibid.
17. D. F. Halpern, *Thought and Knowledge: An Introduction to Critical Thinking,* 4th ed. (Mahwah, NJ: Erlbaum, 2003); M. Patricia King et al., "Impact of Participant and Physician Intervention Preferences on Randomized Trials: A Systematic Review," *Journal of the American Medical Association* 293, no. 9 (2005): 1089–1099.
18. Macias et al., "Preference in Random Assignment," 331–342.
19. Ibid.
20. Ibid.
21. Sameer B. Srivastava, "Network Intervention: Assessing the Effects of Formal Mentoring on Workplace Networks," *Social Forces* 94, no. 1 (2015): 427–452.
22. Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Designs.*
23. Cengiz Erisen, Elif Erisen, and Binnur Ozkececi-Taner, "Research Methods in Political Psychology," *Turkish Studies* 13, no. 1 (2013): 17.
24. Diana Mutz, *Population-Based Survey Experiments* (Princeton, NJ: Princeton University Press, 2011); Steven L. Nock and Thomas M. Guterbock, "Survey Experiments," in *Handbook of Survey Research*, ed. P. V. Marsden and J. D. Wright (Bingley, UK: Emerald Group Publishing Limited, 2010), 837–864.
25. Erin C. Cassese et al., "Socially Mediated Internet Surveys: Recruiting Participants for Online Experiments," *PS: Political Science & Politics* 46, no. 4 (2013): 1–10; James N. Druckman, "Priming the Vote: Campaign Effects in a US Senate Election," *Political Psychology* 25, no. 4 (2004): 577–594; Julio J. Elias, Nicola Lacetera, and Mario Macis, "Markets and Morals: An Experimental Survey Study," *PLoS ONE* 10, no. 6 (2015): 1–13.
26. Erisen, Erisen, and Ozkececi-Taner, "Research Methods in Political Psychology," 17.
27. Scott Clifford and Jennifer Jerit, "Is There a Cost to Convenience? An Experimental Comparison of Data Quality in Laboratory and Online Studies," *Journal of Experimental Political Science* 1, no. 2 (2014): 120–131.
28. Rebecca B. Morton and Kenneth C. Williams, *Experimental Political Science and the Study of Causality: From Nature to the Lab* (New York: Cambridge University Press, 2010).
29. Jona Linde and Barbara Vis, "Do Politicians Take Risks Like the Rest of Us? An Experimental Test of Prospect Theory under MPs," *Political Psychology* 38, no. 1 (2017): 101–117.
30. Clifford and Jerit, "Is There a Cost to Convenience?" 120–131.
31. Cassese et al., "Socially Mediated Internet Surveys"; Kevin J. Mullinix et al., "The Generalizability of Survey Experiments," *Journal of Experimental Political Science* 2 (2015): 109–138.
32. Renita Coleman and Lee Wilkins, "The Moral Development of Journalists: A Comparison With Other Professions and a Model for Predicting High Quality Ethical Reasoning," *Journalism and Mass Communication Quarterly* 81, no. 3 (2004): 511–527.
33. Renita Coleman and Lee Wilkins, "Searching for the Ethical Journalist: An Exploratory Study of the Ethical Development of News Workers," *Journal of Mass Media*

Ethics 17, no. 3 (2002): 209–255; Renita Coleman and Lee Wilkins, "The Moral Development of Journalists"; R. Coleman and L. Wilkins, "The Moral Development of Public Relations Practitioners: A Comparison with Other Professions," *Journal of Public Relations Research* 21, no. 3 (2009): 318–340.
34. Renita Coleman, "The Moral Judgment of Minority Journalists: Evidence from Asian American, Black, and Hispanic Professional Journalists," *Mass Communication and Society* 14, no. 5 (2011): 578–599.
35. Erisen, Erisen, and Ozkececi-Taner, "Research Methods in Political Psychology," 17.
36. Kate E. West, "Who Is Making the Decisions? A Study of Television Journalists, Their Bosses, and Consultant-Based Market Research," *Journal of Broadcasting & Electronic Media* 55, no. 1 (2011): 27.
37. Jun Myers, "Stalking the 'Vividness Effect' in the Preventive Health Message: The Moderating Role of Argument Quality on the Effectiveness of Message Vividness," *Journal of Promotion Management* 20, no. 5 (2014): 634.
38. Mutz and Pemantle, "Standards for Experimental Research," 192–215.
39. Bausell, *Conducting Meaningful Experiments*.
40. Mark L Davison and Stephen Robbins, "The Reliability and Validity of Objective Indices of Moral Development," *Applied Psychological Measurement* 2, no. 3 (1978): 391–403; James R. Rest, *Development in Judging Moral Issues* (Minneapolis, MN: University of Minnesota Press, 1979).
41. Bausell, *Conducting Meaningful Experiments*.
42. Christopher H. Rhoads and Charles Dye, "Optimal Design for Two-Level Random Assignment and Regression Discontinuity Studies," *Journal of Experimental Education* 84, no. 3 (2016): 421–448.
43. Srivastava, "Network Intervention."
44. D. Fergusson et al., "Post-Randomisation Exclusions: The Intention to Treat Principle and Excluding Patients from Analysis," *BMJ* 325 (2002): 652–654.
45. Howard Levene, "Robust Tests for Equality of Variances," in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, ed. Ingram Olkin et al. (Stanford, CA: Stanford University Press, 1960), 278–292.
46. Mutz and Pemantle, "Standards for Experimental Research," 192–215
47. M. J. Strube, "Small Sample Failure of Random Assignment: A Further Examination," *Journal of Consulting & Clinical Psychology* 59, no. 2 (1991): 346–350.
48. Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Designs*
49. John Graham et al., "Random Assignment of Schools to Groups in the Drug Resistance Strategies Rural Project: Some New Methodological Twists," *Prevention Science* 15, no. 4 (2014): 516–525.
50. Ibid.
51. L. M. Hsu, "Random Sampling, Randomization, and Equivalence of Contrasted Groups in Psychotherapy Outcome Research," *Journal of Consulting and Clinical Psychology* 57 (1989): 131–137.
52. Alan Gerber et al., "Reporting Guidelines for Experimental Research: A Report from the Experimental Research Section Standards Committee," ibid.1, no. 1 (2014): 81–98.
53. Myers, "Stalking the 'Vividness Effect,'" 636.
54. Shirley S. Ho and Douglas M. McLeod, "Social-Psychological Influences on Opinion Expression in Face-to-Face and Computer-Mediated Communication," *Communication Research* 35, no. 2 (April 2008): 190–207.
55. Ibid., 197.
56. Gary W. Ritter and Rebecca A. Maynard, "Using the Right Design to Get the 'Wrong' Answer? Results of a Random Assignment Evaluation of a Volunteer Tutoring Programme," *Journal of Children's Services* 3, no. 2 (2008): 4–16.
57. Merton S. Krause and Kenneth I. Howard, "What Random Assignment Does and Does Not Do," *Journal of Clinical Psychology* 59, no. 7 (2003): 751–766.
58. Ibid.
59. Bausell, *Conducting Meaningful Experiments*; Hsu, "Random Sampling"; Strube, "Small Sample Failure of Random Assignment," 346–350.
60. James R. Rest et al., *Postconventional Moral Thinking: A Neo-Kohlbergian Approach* (Mahwah, NJ: Erlbaum, 1999).
61. Krause and Howard, "What Random Assignment Does and Does Not Do."
62. Mutz and Pemantle, "Standards for Experimental Research."
63. Ibid.
64. Ibid.

65. Hsu, "Random Sampling."
66. Strube, "Small Sample Failure of Random Assignment."
67. Graham et al., "Random Assignment of Schools."
68. Hsu, "Random Sampling."
69. Strube, "Small Sample Failure of Random Assignment."
70. Bausell, *Conducting Meaningful Experiments.*
71. Ibid.
72. Donald P. Green and Daniel Winik, "Using Random Judge Assignments to Estimate the Effects of Incarceration and Probation on Recidivism among Drug Offenders," *Criminology* 48, no. 2 (2010): 357–387.
73. Mutz and Pemantle, "Standards for Experimental Research."
74. Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Designs.*
75. Ibid.
76. Strube, "Small Sample Failure of Random Assignment," 346–350.
77. Krause and Howard, "What Random Assignment Does and Does Not Do"; Strube, "Small Sample Failure of Random Assignment."
78. Graham et al., "Random Assignment of Schools."
79. Campbell and Stanley, *Experimental and Quasi-Experimental Designs for Research.*
80. Murray Webster and Jane Sell, *Laboratory Experiments in the Social Sciences* (Amsterdam: Elsevier, 2007).
81. Bausell, *Conducting Meaningful Experiments*; Campbell and Stanley, *Experimental and Quasi-Experimental Designs for Research.*
82. Srivastava, "Network Intervention."
83. Ibid., 438.
84. Graham et al., "Random Assignment of Schools."
85. David Alan Free, "Perpetuating Stereotypes in Television News: The Influence of Interracial Contact on Content" (unpublished dissertation, University of Texas, 2012).
86. E. W. Gondolf, "Lessons from a Successful and Failed Random Assignment Testing Batterer Program Innovation," *Journal of Experimental Criminology* 6 (2010): 355–376.
87. Mutz and Pemantle, "Standards for Experimental Research."
88. Campbell and Stanley, *Experimental and Quasi-Experimental Designs for Research.*
89. Ruxton and Colegrave, *Experimental Design for the Life Sciences.*
90. For example, see ibid.; Murray R. Selwyn, *Principles of Experimental Design for the Life Sciences* (Boca Raton, FL: CRC Press, 1996).
91. Campbell and Stanley, *Experimental and Quasi-Experimental Designs for Research*; Strube, "Small Sample Failure of Random Assignment."
92. Krause and Howard, "What Random Assignment Does and Does Not Do."
93. Campbell and Stanley, *Experimental and Quasi-Experimental Designs for Research,* 15.
94. Selwyn, *Principles of Experimental Design for the Life Sciences.*
95. Graham et al., "Random Assignment of Schools."
96. Ibid.
97. Ibid.
98. Ibid.
99. Bausell, *Conducting Meaningful Experiments.*
100. Graham et al., "Random Assignment of Schools," 521.
101. Ibid.
102. Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Designs,* 253.
103. Bausell, *Conducting Meaningful Experiments*.
104. Linde and Vis, "Do Politicians Take Risks Like the Rest of Us?"
105. Ibid., 107–108.
106. Ibid., 111.
107. Bausell, *Conducting Meaningful Experiments,* 78.
108. Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Designs.*
109. Bausell, *Conducting Meaningful Experiments*.
110. Roger Kirk, *Experimental Design: Procedures for the Behavioral Sciences,* 4th ed. (Los Angeles, CA: Sage, 2013).
111. Campbell and Stanley, *Experimental and Quasi-Experimental Designs for Research.*
112. Ruxton and Colegrave, *Experimental Design for the Life Sciences.*
113. Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Designs*.
114. Kirk, *Experimental Design.*
115. Zhang Jueman, Zhang Di, and T. Makana Chock, "Effects of HIV/AIDS Public Service Announcements on Attitude and Behavior: Interplay of Perceived Threat

and Self-Efficacy," *Social Behavior and Personality: An International Journal* 42, no. 5 (2014): 799–809.
116. West, "Who Is Making the Decisions?" 27.
117. Eron M. Berg and Louis G. Lippman, "Does Humor in Radio Advertising Affect Recognition of Novel Product Brand Names?" *Journal of General Psychology* 128, no. 2 (2001): 194.
118. Ibid., 199.
119. Kirk, *Experimental Design*, 671.
120. E. L. Thorndike, W. A. McCall, and J. C. Chapman, *Ventilation in Relation to Mental Work*, vol. 78, Teachers College Contribution to Education (New York: Teachers College, Columbia University, 1916).
121. Ronald A. Fisher, *Statistical Methods for Research Workers* (Edinburgh: Oliver and Boyd, 1925).
122. Campbell and Stanley, *Experimental and Quasi-Experimental Designs for Research.*
123. Ritter and Maynard, "Using the Right Design to Get the 'Wrong' Answer?"
124. Ibid.
125. Srivastava, "Network Intervention."
126. Yun Hyung Koog and Byung-Il Min, "Does Random Participant Assignment Cause Fewer Benefits in Research Participants? Systematic Review of Partially Randomized Acupuncture Trials," *Journal of Alternative and Complementary Medicine* 15, no. 10 (2009): 1107–1113.
127. Richard A. Berk, Gordon K. Smyth, and Lawrence W. Sherman, "When Random Assignment Fails: Some Lessons from the Minneapolis Spouse Abuse Experiment," *Journal of Quantitative Criminology* 4, no. 3 (1988): 209–223.
128. Ong-Dean, Huie Hofstetter, and Strick. "Challenges and Dilemmas."
129. Gueron, "The Politics of Random Assignment," 26.
130. Ong-Dean, Huie Hofstetter, and Strick, "Challenges and Dilemmas," 39.
131. Ibid., 45.
132. Gondolf, "Lessons from a Successful and Failed Random Assignment."
133. Ong-Dean, Huie Hofstetter, and Strick, "Challenges and Dilemmas."
134. Gary W. Ritter and Marc J. Holley, "Lessons for Conducting Random Assignment in Schools," *Journal of Children's Services* 3, no. 2 (2008): 28–39.
135. Ibid.
136. Ibid.
137. Ibid.
138. Ibid.
139. Ritter and Maynard, "Using the Right Design."
140. Ritter and Holley, "Lessons for Conducting Random Assignment in Schools."
141. Ibid.
142. Bausell, *Conducting Meaningful Experiments*; Ritter and Holley, "Lessons for Conducting Random Assignment in Schools."
143. Ibid.
144. Ibid.
145. Ong-Dean, Huie Hofstetter, and Strick, "Challenges and Dilemmas"; Ritter and Holley, "Lessons for Conducting Random Assignment in Schools."
146. Ong-Dean, Huie Hofstetter, and Strick, "Challenges and Dilemmas"; Ritter and Holley, "Lessons for Conducting Random Assignment in Schools."
147. Berk, "Randomized Experiments as the Bronze Standard"; Ong-Dean, Huie Hofstetter, and Strick, "Challenges and Dilemmas."
148. Bausell, *Conducting Meaningful Experiments*
149. Ibid.
150. Ritter and Holley, "Lessons for Conducting Random Assignment in Schools."
151. Ibid.
152. Ong-Dean, Huie Hofstetter, and Strick, "Challenges and Dilemmas."
153. Ritter and Holley, "Lessons for Conducting Random Assignment in Schools."
154. Ong-Dean, Huie Hofstetter, and Strick, "Challenges and Dilemmas."
155. Ibid.
156. Srivastava, "Network Intervention."
157. Gondolf, "Lessons from a Successful and Failed Random Assignment."
158. Ritter and Holley, "Lessons for Conducting Random Assignment in Schools."
159. Bausell, *Conducting Meaningful Experiments.*
160. Ong-Dean, Huie Hofstetter, and Strick, "Challenges and Dilemmas."
161. Ibid.
162. Ibid.
163. Ibid.
164. Ibid.

165. Ritter and Holley, "Lessons for Conducting Random Assignment in Schools."
166. Ibid.
167. Ong-Dean, Huie Hofstetter, and Strick, "Challenges and Dilemmas."
168. Ibid.; Ritter and Holley, "Lessons for Conducting Random Assignment in Schools."
169. Berk, Smyth, and Sherman, "When Random Assignment Fails."
170. Ong-Dean, Huie Hofstetter, and Strick, "Challenges and Dilemmas"; Koog and Min.
171. Berk, Smyth, and Sherman, "When Random Assignment Fails."
172. Bausell, *Conducting Meaningful Experiments*, 74.
173. Gondolf, "Lessons from a Successful and Failed Random Assignment."
174. Ong-Dean, Huie Hofstetter, and Strick, "Challenges and Dilemmas."
175. Ibid.
176. Ritter and Holley, "Lessons for Conducting Random Assignment in Schools."
177. Ibid.
178. Ibid.
179. Ibid.
180. Campbell and Stanley, *Experimental and Quasi-Experimental Designs for Research*, 15.
181. David S. Abrams and Albert H. Yoon, "The Luck of the Draw: Using Random Case Assignment to Investigate Attorney Ability," *University of Chicago Law Review* 74, no. 4 (2007): 1145–1177.