Source: iStock.com/gorodenkoff

2 DOING HEALTH PSYCHOLOGY

Research Methods

LEARNING OBJECTIVES

2.1 Compare and contrast the major research designs.

2.2 Describe the key statistical considerations in analyses.

2.3 Explain the difference between different forms of variables.

2.4 Identify important considerations in doing research on health and culture.

"Coffee is good for you." "Coffee is bad for you." "Eating chocolate leads to a longer life." "Cell phone usage may cause cancer." "Daily mindfulness interventions lead to happiness." You have probably seen social media blaring similar headlines. Our Facebook and Twitter feeds are often flooded with the latest "Research shows . . . " type shares. Often, the advice is contradictory. Sometimes what is good for us in one year is bad for us in another year. Whereas it is easy to think nothing is true and that health is too complex to fully predict, the reality is that the media do not always do a good job of reporting research. We textbook authors aim to do a much better job but you will be a better consumer of information (both in life and in understanding health psychology) if you have a working understanding of the basic elements of research, the major designs used, and some common statistical analyses (and their interpretation). Scientific knowledge and research has a toolkit and a common methodology. This textbook is based on peer-reviewed journal articles—research published in academic journals that have passed the tests of independent review. How is that research done? What are the major designs used? In this chapter, I will give you the tools to enable you to open up any research journal and be able to better understand the findings discussed within. By chapter's end we get to a major question for all scientific research—Do the findings replicate?

PONDER THIS

Before you share a research study on Facebook, Twitter, or other social media, what should you be looking for in order to know it is valid?

What are the different ways to design research?

Does a startling research finding need to be shown again (replicate) before you believe it? Why or why not?

MEASURING UP: IS PSYCHOLOGY A SCIENCE?

Listed below are a number of statements. Read each statement carefully and indicate the extent to which you agree or disagree by writing the appropriate number by each statement on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*).

1. An undergraduate degree in psychology should be a bachelor of science rather than a bachelor of arts degree. ________
2. It's just as important for psychology students to do experiments as it is for students in chemistry and biology. ________
3. Research conducted in controlled laboratory settings is essential for understanding everyday behavior. ________
4. Even though each person is unique, it is possible for science to find general laws explaining human behavior. ________
5. Carefully controlled research is not likely to be useful in solving psychological problems. ________
6. Our ability as humans to behave in any way we choose makes our attempts to predict behavior ineffective. ________
7. Psychological advice given in popular books and magazines is often as useful as claims that are more research based. ________
8. Government funding of experimentation is as necessary for expanding what we know about psychology as it is for gaining knowledge in areas like chemistry and physics. ________
9. The study of psychology should be seen primarily as a science. ________
10. Courses in psychology place too much emphasis on research and experimentation. ________
11. Psychological research can enable us to anticipate people's behavior with a high degree of accuracy. ________
12. Psychologists working as counseling professionals don't need to be so concerned with research findings. ________
13. Psychological theories presented in the media should not be trusted unless they are supported by experiments. ________

14. Psychology will never be a true science because its predictions of individual behavior are seldom exact or certain. ________

15. Students get little benefit from learning about procedures for conducting psychology experiments. ________

Source: "The Psychology as a Science Questionnaire" (Friedrich, 1996).

Regardless of which definition of health we consider, each definition of health is broad and ambiguous. How can we measure mental, spiritual, and social health? Does simply the absence of physical problems or disease equate to health? Can anyone even measure a balanced yin and yang? The answer is no, not really, or at least not by any measure that we know of or use in the United States or in the scientific community, and not in a way on which we can all agree. To understand what keeps us healthy, it is important to start with a good measurement of health. As you learn about the field of health psychology, you will see that although most researchers will use a common understanding and relatively broad definition of health to guide their general thinking (e.g., a general state of well-being), every researcher uses a different specific measure of health to help understand what makes us healthy.

Take a quick look at the major research journals that report on health psychological research, and you will see that different studies use slightly different measures. This is the first major element to watch for when reading articles. The main categories of measures vary with each journal. For example, *Health Psychology* is the leading journal in the field and publishes the results of studies on the topic of health psychology. This journal features many studies that define health in terms of the extent to which health-improving behaviors are practiced (e.g., how much did the participants in the study exercise in a week?) or in terms of psychological well-being (e.g., what were the participants' scores on the Profile of Mood States, a common measure of mood?). You will also see many studies that assess the extent to which health-diminishing behaviors are practiced. For example, how much does a person smoke? What predicts the amount of alcohol consumed?

Other journals, such as the *Annals of the Society for Behavioral Medicine* and *Psychosomatic Medicine*, measure many specific physiological outcomes. For example, what are the levels of immune cells in the blood? Figure 2.1 shows sample contents from the three major journals. The bottom line is that we determine if people are healthy by measuring a variety of aspects. You will see measures of basic physiological

FIGURE 2.1 ■ **Journal Covers**

Sources: Applied Psychological Measurement, 42(3); *Cross-Cultural Research, 52*(2); *Clinical Case Studies, 17*(2).

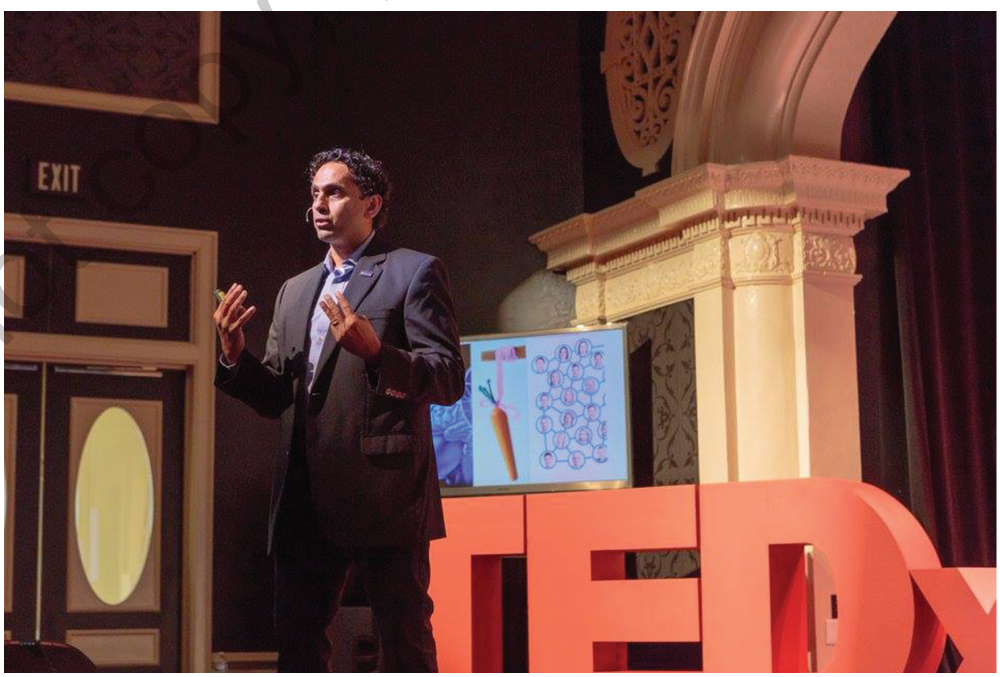
levels of bodies' various systems (e.g., blood pressure, heart rate, or cholesterol level). You will see measures of how much people practice healthy behaviors (e.g., exercising). You will also see many measures of psychological well-being (e.g., levels of depression or optimism) and how well people practice healthy psychological ways (e.g., good coping skills). There are so many different ways to measure the key elements of health, an entire book focuses on the different measures used in health psychology (Benyamini et al., 2016).

A RESEARCH PRIMER

Health psychology relies firmly on the scientific method. The key elements of a science are (1) that it is empirical (relying on sense observations and data) and (2) that it is theory driven. The data or empirical evidence is collected in ethical, rigorously controlled, and standardized ways whether you are identifying causes of stress or testing the psychological effects of an intervention to reduce smoking. The research enterprise is a fascinating one; to get a good feel for the results of research (discussed throughout this book), you should have a good idea of the main research designs and data collection methods. Because the bulk of our knowledge comes from research, courses in experimental methods and statistics are great companions or foundations for the health psychology course. This chapter should be a good refresher for those of you who have taken such courses or provide the rest of you with enough to really enjoy reading research journals.

Understanding the common research methods used by health psychologists and knowing how to interpret common statistical results will also enable you to make better sense of peer-reviewed journal articles, the source of the information used in writing this book. Even if you learn of results of research on the radio, television, or via the internet, it is always good practice to go to the original published article to substantiate the results. You will be surprised how often media outlets spin a finding to amplify the possible implications. Reading the original sources for yourself (and understanding them) will make you a better consumer of science. You may want to head over to PsycInfo (or a related database) right now and look up the latest issue of *Health Psychology*. Then go back to it after you read the rest of this chapter and feel the thrill of being research savvy.

Watch for different ways studies are set up. There are a wide variety of research designs in health psychology (Lovejoy & Fowler, 2019). As discussed in Chapter 1, Health Psychologists may be trained



Research Presentations. We often hear about research from conference presentations or even listening to TED talks. Listen to mine on how to Chill, Drill, and Build for healthy living. Remember, neither go through the same level of peer review as do scholarly journals.

Dr. Regan A. R. Gurung

as clinicians, experimentalists, developmentalists, or social psychologists. Health psychology research is also conducted in different health-related disciplines such as nursing, medicine, and public health. Each of these different areas favors different research designs. Someone with a public health degree may favor population-wide measurement of bike path usage. Developmental psychologists may prefer studies conducted over a period of time. Social psychologists may prefer experimental tests of different treatments. Each design has its pros and cons, as you will see in the pages ahead.

Key Steps to Doing Research

The steps for conducting research on health mirror most of the steps used to conduct research on any topic. First, the researcher identifies a question of interest and then reviews what has been published on the topic. Next, the researcher ascertains what is left to be discovered or needed to be researched and then decides how to conduct that research. One common approach to this sort of research is to measure basic relevant health behaviors or states of health whether psychological or physiological. There are a variety of specific scales or questionnaires one can use to measure, assess, or evaluate health (Luszczynska et al., 2019). *Measurement* focuses on describing characteristics of an individual (e.g., self-efficacy). *Assessment* relates to obtaining information according to a goal (e.g., how much did exercise increase after an intervention). *Evaluation* "accounts for the individual in a specific situation and the goals or criteria which were externally set resulting in a normative judgment" (Luszczynska et al., 2019). In health psychology, these terms are often used interchangeably.

Similar to conducting psychological science in general, health researchers can start with descriptive studies (e.g., How much are people exercising?), move on to correlational designs (e.g., What is exercising associated with?), and then design interventions (e.g., If I introduce a new way of talking about exercise, will amount of exercise change?). In designing interventions, researchers can choose from a wide variety of models. Lovejoy and Fowler (2019) present a fuller description of different research designs and particular adaptations for health psychological interventions.

Good news. The basic process of doing research is common across the board and a recent chapter in the *Handbook of Health Psychology* (Revenson & Gurung, 2019) provides us with major research designs in health psychology (Ranby, 2019). I summarize the main steps in Table 2.1. As you can see, most map nicely onto what you have read about in your research methods class. There are some nuances, of course. In health psychology research, you tend to have to spend more time and energy collecting data and are more likely to be measuring sensitive topics. You have to pay particular attention to selecting your measurement tools, as well. Let's first overview the major types of research designs used in health psychology. We can then focus on measurement.

TABLE 2.1 ■ Major Steps in Research Design

1. Plan Your Study **a.** Consider a program of research that plans, in advance, for different studies to answer the key questions. **b.** Clarify the study's main purpose. **c.** Identify a target population. **d.** Consider if important subgroups exist in the population. **e.** Predict or hypothesize the effect you expect.
2. Pick a Design **a.** Descriptive **b.** Correlational **c.** Experimental **d.** Longitudinal
3. Recruit Participants
4. Conduct study
5. Analyze Data
6. Replicate and Report

Source: Adapted from Ranby (2019)

RESEARCHER SPOTLIGHT

Dr. Krista Ranby earned her PhD in social psychology from Arizona State University, Tempe. She currently teaches at the University of Colorado, Denver, and is skilled at research design. See her chapter in Essential Readings (Ranby, 2019).

Major Research Designs

Descriptive Studies

The most basic form of research describes what is going on: How many people smoke? How prevalent is a certain disease or disorder? If you receive an electronic survey over email asking you about your behaviors you are probably being recruited for a descriptive study. This basic form of design is exploratory and aims to establish baselines for behaviors. You are most likely to see descriptive studies in the field of epidemiology and public health. Epidemiological studies often report **prevalence rates**, the proportion of the population that has a particular disease at a particular time (commonly reported as cases per 1,000 or 100,000 people), and **incidence rates**, the frequency of new cases of the disease during a year.

We see descriptive studies all around us. Take this headline: "Too many older patients get cancer screenings" (Szabo, 2017). This particular *New York Times* article reports on a study published in the *American Journal of Public Health* (Mehta et al., 2017) showing that nearly one in five older women are getting regular mammograms when they should not. The issue is that mammograms are not recommended for people with limited life expectancy. The study does not tell you why or if there is a downside, but just reports on the numbers of individuals doing the certain behavior. Make sure you do not draw conclusions beyond what is measured in the study. Well-written research articles do not go beyond the data but we cannot expect the same from media reports of those same articles. The more research savvy you are the less likely you are to waste time and energy on false conclusions.

Correlational Studies

The most basic form of research design describes relationships between variables. Are heavier people at more risk for cardiovascular disease? Do poorer people smoke more? A **correlation coefficient** is the statistical measure of the association represented with a lower case *r*. Correlations range from –1.00 to +1.00 with values closer to 1 (regardless of sign) signifying stronger associations. This is a key point because $r = -.54$ is stronger than $r = .38$. Many novices see the negative sign and assume it is not good. Wrong. The sign refers to the direction of the relationship. Positive correlations indicate variables that change in the same direction (e.g., higher weight correlates with higher risk of cardiovascular disease). Negative correlations indicate variables that change in opposite directions (e.g., lower **socioeconomic status** correlates with higher smoking rates).

As you read journal articles, do not be surprised to see emphasis about correlations around the .2 to .3 level. If 1.00 is the highest, .2 and .3 seem low. They are, but given that there are many factors accounting for any behavior or result, a *statistically significant* correlation in the .2 to .3 range between any two variables suggests a worthy relationship. Make sure you also pay attention to statistical significance. The results section of a journal article will report the significance level of any findings. Statistical analyses with *p*-values (probability values) less than. 05, .01, or .001 (reported as p <.05, p <.01, or p <.001) are significant. All results tables will indicate statistically significant differences with an asterisk (*) so you can be on the lookout for these asterisks. The more asterisks you see, the higher the statistical significance (and unlike correlations, the *lower* the *p* value, the better). A probability value of less than 0.05 suggests that the probability of getting the same result by chance is less than 5 in 100. You can see why $p < 0.001$ is a significant level. The *p* value is influenced by sample size; given that many studies in health psychology have very large samples, you will note that even small correlations could be statistically significant.

Many journal articles in health psychology will report correlations between variables although I have noticed that more-complex statistics (such as odds ratios) are replacing this simple statistic. Speaking of simple, when only the relationship between two variables is tested (e.g., distress correlated with coping style) you call it a zero-order or direct correlation. Yes, this is helpful and illustrates a relationship, but quite honestly it is somewhat misleading. There is often more than one variable influencing another; to statistically control for multiple associations, researchers use a **partial correlation**. When you calculate a partial correlation or control for another variable, the relationship between two variables is tested while controlling for a third variable (or more). For example, researchers often statistically control for a research participant's age when assessing correlations, which essentially acknowledges that the association between the variables of interest (e.g., distress and coping style) could vary for people of different ages.

When you read a journal article, do not be intimidated by the large mass of numbers you see. Even correlational tables may seem like a mess of numbers, but take your time to orient yourself to what you are seeing. Take a look at the correlational table presented in Table 2.2. In this study, researchers wanted to test the link between factors such as ethnic discrimination and risky behaviors (Bravo et al., 2017). All the participants were Mexican-origin adolescents. Although there are a lot of numbers, orient yourself to the basic elements of the American Psychological Association style table.

The main variables are listed in the column down the left (e.g., 1. Age, 2. Nativity). The same variables are referred to along the top, and, conveniently, the labels (e.g., Age) do not have to be rewritten. Each dash (–) represents the association of a variable with itself. Given that one variable will be perfectly

TABLE 2.2 ■ Means, Standard Deviations, and Correlations Among Study Variables for Sample (*N* = 204)

Study variables	1	2	3	4	5	6	7	8	9	10	11	12
1. W1 age	—											
2. W1 nativity[a]	.06	—										
3. W4 economic hardship	–.05	.18**	—									
4. W4 ethnic discrimination	–.01	.13	.24***	—								
5. W4 family support	.07	–.06	–.19**	–.19**	—							
6. W4 friend support	.03	–.03	–.14*	–.16*	.37***	—						
7. W4 risky behaviors	–.04	.16*	.18**	.27***	–.19**	–.11	—					
8. W5 family support	.05	–.09	–.19**	–.22**	.64***	.25***	–.19**	—				
9. W5 family support	.18**	.04	–.20**	–.14*	.28***	.63***	–.08	.40***	—			
10. W5 risky behaviors	.01	.13	.20**	.28***	–.30***	–.06	.06***	–.36***	–.05	—		
11. W6 body mass index	–.10	.06	.10	–.13	–.03	.00	.11	–.04	–.02	.13	—	
12. W2 to W6 pregnancy status[b]	–.08	–.04	.08	.09	–.09	–.15*	–.10	–.04	–.17*	–.06	–.01	—
Mean	16.81	.64	.01	1.32	5.90	5.50	1.24	5.92	5.53	1.24	26.71	.55
Standard Deviation	.99	.48	2.37	.42	1.17	1.34	.30	1.21	1.27	.28	5.26	.50

Source: Bravo, D. Y., Derlan, C. L., Umaña-Taylor, A. J., Updegraff, K. A., & Jahromi, L. B. (2017). Processes underlying Mexican-origin adolescent mothers' BMI. *Cultural Diversity and Ethnic Minority Psychology*. Reprinted with permission of APA.

Note: W = Wave. Means and Standard deviations reported here were calculated prior to centering. W4 economic hardship is calculated as a weighted summed score.

[a]Nativity coded as: 0 = Mexico born; 1 = U.S. born. [b]Pregnancy status was coded as 0 = *not pregnant after W1* or 1 = *pregnant at least once between W2 to W6*.

*p <.05. **p <.01. ***p <.001.

(i.e., 1) correlated with itself, the numeral 1 is not used. All the numbers with asterisks represent the statistically significant correlations. The more *s, the higher the statistical significance. Note that body mass index, the major medical measure used to assess obesity, is not related to any other measures. None of the numbers in row 11 have any asterisks. You can see negative correlations (e.g., nativity and economic hardship). The U.S.-born participants were less likely to have economic hardship. There are some positive correlations too. The more economic hardship reported by participants, the more ethnic discrimination they report ($r = .24$). Once your eyes orient to the numbers there is a lot of information to pull from this one table.

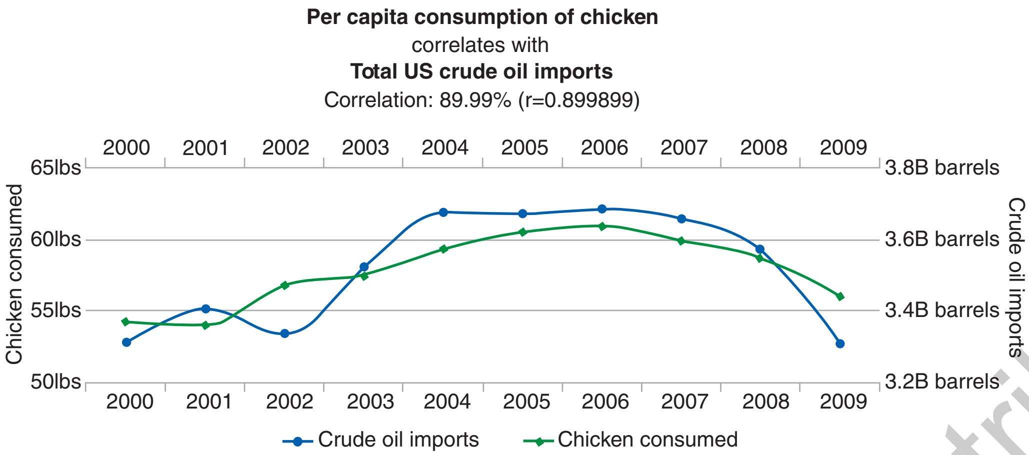
There are many examples of **correlational studies** around. Something that the lay public often ignores is that many of the studies on issues they care about are correlational studies. A common interest? What to eat. Most food studies, especially those with large sample sizes, are correlational. For example, Dehghan et al. (2017) conducted a large epidemiological study in 18 countries. They measured the diets of 134,335 individuals and followed them over a 10-year period. At the end of that time, the researcher measured how many people were alive, and how many had had heart attacks. In the latter group, they looked to see if the deaths were linked to their diets. Higher-carbohydrate intake was linked to higher deaths but not with risk for heart attacks. [A geeky aside: I cannot help but share that this article had more than 400 authors. Probably the best use of the et al. citation style ever. Lucky Mahshid Dehghan who was first author and the only one you see unless you go to the references at the end of the book.]

Correlational designs do not allow us to draw causal conclusions. In the study above can we conclude eating carbs kills? Is this reason enough to start piling on the bacon? Not exactly. This does not stop the casual reader or the uninformed, and can sometimes lead to mild panic where none is warranted. For example, another study found that oral sex was correlated with incidences of mouth cancer (Kreimer et al., 2004). Does this mean that practicing oral sex causes mouth cancer? If the study was correlational, it does not. In fact, the study also showed that oral sex was highly correlated with smoking and drinking, two behaviors more likely to cause mouth cancer. See Figure 2.2 for how correlations can be misleading.

What type of food is best for you? Research on carbs, proteins, and fats is heavily debated, an especially good reason to know what makes good research.

Source: iStock.com/filadendron

FIGURE 2.2 ■ Correlations. There are many variables that seem to be associated statistically that are clearly not causing each other.



Source: Per capita consumption of chicken correlates with total U.S. crude oil imports by Tyler Vigen, https://tylervigen.com/spurious-correlations, licensed under CC BY 4.0 https://creativecommons.org/licenses/by/4.0/.

Experimental and Quasi-Experimental Designs

Experimental designs help us determine causality. All health psychology interventions are, by definition, experiments. If we really want to know if something caused something else we need to introduce that something and see if it has an effect. If carbohydrates are bad for health, we need to design a study that measures health first, then introduces carbohydrates, and then measures health afterward, controlling for other factors. If health changes after the carbohydrates are introduced and nothing else changes, then we can conclude carbohydrates cause poor health. In experiments, the researcher manipulates the variable that is believed to be important—the **independent variable**—and measures how changes in this variable influence another variable—the **dependent variable**.

Experiments have two or more groups, each of which experiences different levels of the independent variable. If the participants are randomly sampled (everyone in the population has an equal chance of being in the study) and extraneous variables (other variables that may influence the outcome of interest such as socioeconomic status or other health behaviors) are controlled for, then one can be fairly certain that changes in the dependent variable are due to changes in the independent variable. Cause can be determined. Many of the studies in this book are experiments.

To test whether exercise is good for concentration, you can have one group of people exercise (the independent variable) three times a week and the other not exercise. You can then see if the two groups vary in concentration (the dependent variable). In health psychological research, it is often impractical and unethical to manipulate key variables of interest (e.g., making people smoke or have oral sex) so groups that naturally vary in the variable of interest are used instead (e.g., compare groups of people who vary in how much they smoke or have oral sex). Because using naturally occurring groups is not a perfect experiment, such designs are referred to as **quasi-experimental designs**, and the independent variables are called subject variables. Examples of common subject variables are age, sex, ethnicity, personality type, occupation, socioeconomic status, and disease state (level or presence of).

Whereas in correlations you look at how close to 1 the r value is and whether it is statistically significant, in experiments you look at whether the test for group differences is statistically significant. The main statistics you look for a *F tests* or *t tests* (more on this later in this chapter).

Randomized Control Trials

In the health psychology world, experiments more often take the form of **randomized, controlled, or clinical trials (RCTs)**, in which one group gets an experimental drug or intervention treatment and a second group unknowingly gets a **placebo** (an inactive substance that appears similar to the experimental drug) or nothing (the control group). A large number of evidence-based treatment reviews and clinical interventions use RCTs.

Most RCTs are to test interventions and use clinical trials that begin with controlled study designs, have restricted patient samples, and include well-trained researchers to maximize internal validity. Internal validity essentially means ensuring that the active intervention, and not other factors, caused observed changes in the outcome (Lovejoy & Fowler, 2019). These trials are referred to as clinical *efficacy* (or Phase III) trials. There are also clinical *effectiveness* (or Phase IV) trials where an intervention is delivered with the goal of testing how well it will generalize to a large sample (i.e., building external validity).

A useful heuristic for developing health psychology interventions for chronic diseases was recently proposed by the Obesity-Related Behavioral Intervention Trials (ORBIT) consortium, in conjunction with National Institute of Health (NIH) representatives and other experts in health-related behavioral treatments (Czajkowski et al., 2015). According to the ORBIT model, the process begins with identification of a key clinical problem that is catalyzed or perpetuated by a behavioral, psychological, and/or social factor, and that could thus be remediated with a health psychology intervention. In the subsequent design phase, researchers conduct systematic reviews, meta-analyses, epidemiologic research, small-sample experimental studies, and qualitative research to establish evidence for the pathway between a behavioral, psychological, or social risk factor and a meaningful clinical or biological outcome (Lovejoy & Fowler, 2019). Preliminary testing follows the design phase.

Perhaps one of the best examples of an RCT is the Women's Health Initiative study that was launched in 1991 and in which more than 161,000 healthy postmenopausal women were given hormone replacement pills or a placebo. Researchers stopped the study before completion because the results indicated that women taking the pills were actually more at risk for heart disease (Manson et al., 2003). To make matters worse, recent results suggest that even after stopping the study, women who received the hormone replacement pills still had a higher risk of heart disease (Heiss et al., 2008). The data set is so rich and detailed that researchers continue to mine it to answer questions about a range of illnesses and predictors of mortality (Chen et al., 2017; Jones et al., 2017).

Cross-Sectional and Longitudinal Designs

Research can also be **cross-sectional**, conducted at one point in time, or **longitudinal**, conducted over a period of time and often involving many measures of the key variables. Cross-sectional studies often sample a large number of people and examine different cultural groups in the sample comparing men and women, and people of different ethnicities.

Research can be **prospective**, following disease-free participants over a period to determine whether certain variables (e.g., eating too much fast food) predict disease, or **retrospective**, studying participants with a disease and tracing their histories of health behaviors to determine what caused the disease.

There are a number of well-known prospective studies. One study is the Women's Health Initiative described previously. Another study that you will see many references to in the media and in health psychology research is the Nurses' Health Study (NHS). Started in 1976, the NHS and the NHS II are among the largest prospective studies of the risk factors for major chronic diseases in women (e.g., Tamimi et al., 2005). Approximately 122,000 registered nurses in 11 states were followed over time; findings shed light on a variety of health issues ranging from preventing premenopausal colorectal cancer and breast cancer, to the impact of weight on cancer risk (Dworetzky et al., 2012).

ENSURING STRONG MEASUREMENT

Regardless of your research design, you need to have robust measurement. If your measurements are inaccurate or not precise, your results are not valid. There are number of key factors to keep in mind. Health psychology measurement involves five major steps (Luszczynska et al., 2019). These steps involve (1) the choice of a general framework, considering purpose and domains of measurement, (2) characteristics of the target population, (3) the type of measurement, (4) psychometric characteristics, and (5) issues of implementation of an instrument.

It is particularly important to have a general framework that influences either the purpose of the measurement or the domains measured (Karademas et al., 2016). Frameworks relating to the purpose of measurement distinguish between measurement conducted to reach a clinical decision, describe a target population, or clarify health-related processes that predict psychosocial outcomes. Frameworks organizing measurement by the domains in health psychology include domains of health and prevention, stress and health, and illness and care (Luszczynska et al., 2019).

When considering who you want to study, researchers need to plan differently for different age groups as well as different cultural groups in the population. Many measures were created in English but may need to be translated for use with non-English-speaking participants. The cultural adaptation of measures is an important endeavor when one takes a cultural approach to health (López-Roig & Pastor, 2016). Fortunately, there are many guidelines, methods, and procedures for translating measures and working with different cultural populations.

Once you are ready to measure you can pick from one of a variety of measures. Most common are self-reports (diaries, surveys, questionnaires). You can also use biomarker-based measurement (e.g., cortisol in the blood) and one of a wide variety of biological and physiological measures (Segerstrom et al., 2016). With advances in technological innovation you can also use wearable cameras, accelerometers for motion, or sensor-based measurement (e.g., medication events monitoring systems, or MEMS). MEMS include adherence monitoring devices such as electronic pill containers that register and code information about when each pill is taken (Lam & Fresco, 2015).

Need a blast from your (research methods class) past? When we talk about measurement we need to keep in mind the main qualities of measurement. Key psychometric properties include the reliability of the instrument (e.g., its internal consistency, inter-rater reliability, test-retest reliability), its validity (e.g., construct, criterion, concurrent, predictive, or convergent validity), and sensitivity, that is the extent to which the instrument may detect small changes occurring over time (cf. Johnson et al., 2016).

Finally, one needs to measure implementation. How acceptable is your measurement? To what extent were your measurement tools adopted by those who are using them? Besides assessment of implementation of the actions of health psychologists, the other issue refers to the implementation of any measurement instruments, conducted by health psychologists. Would using a certain measure make participants less likely to take part in the study or perhaps to drop out at a later stage? Measurement in health psychology research is not as straightforward as one may think (Luszczynska et al., 2019).

Data Collection. New models of data collection involve the use of smartphones.

Source: iStock.com/DisobeyArt

Getting Statistically Savvy

Together with looking at correlation coefficients (reported using the italicized letter *r*) and tests of group differences (reported using the italicized letters *F* or *t*), there are a number of statistical elements to watch for that can make journal article reading palatable.

Be aware: Not all associations or changes may be statistically significant. Furthermore, not all statistically significant change may be *meaningful* change. We can launch a philosophical debate around the topic of what constitutes meaningful change in weight or happiness. However, there is one simple answer that is hard to negate: statistically significant changes that could not have taken place by chance are important. That said, there are some simple factors that can artificially create statistical significance. The most critical to consider is the number of participants being studied (or the sample size). Researchers perform many health psychological studies on hundreds or thousands of participants. Increasing the sample size can make previously insignificant changes significant.

There are some safeguards and limiting factors. For example, only phenomena that have a large *effect size* will be significant when the sample size increases. If the psychological intervention or the drug tested or a cognitive behavioral change was not effectual (a simple paraphrase of effect size that adequately conveys its intended meaning), many more participants may not make results significant. Most journal articles report effect size, something that is now required if you are writing in American Psychological Association (APA) style. Look for it as one of many Greek letters representing a variety of effect size calculations, most commonly reported using ηp^2 and read as "partial eta squared."

Common Statistical Tests

There a number of major statistical tests you will encounter over and over again. The first time you run into them they may sound like gibberish, but once you walk through it you will see that they are actually quite simple. Let's look at three of them: **analyses of variance (ANOVAs)**, **multivariate analyses of variance (MANOVAs)**, and **regression analyses**.

Both ANOVAs and MANOVAs test for differences between group means. Is the weight loss in one group different from the weight loss in another group (ANOVAs)? If you want to test for differences between a number of variables that are related to each other, you would use a MANOVA (hence, the multivariate). Are the ratings of quality and taste of one type of food better than another (MANOVA)? The good news is that just like for a correlation, you are paying attention to the *p*-value of the statistical test. Look for if it is significant. If the *p*-value of the *F* test, the test used for ANOVA, is less than .05 the means you are comparing are significantly different. When I teach this class, I have all my students complete a short set of questions about their health behaviors and their health. The students then get the data (with names removed, of course) and calculate some simple analyses themselves. They often find some interesting findings. Recently, my class compared the body mass index (BMI) scores of men and women using an ANOVA and found a statistically significant difference where the average BMI scores of women were higher than that for men, $F(1,44) = 5.34$, $p = .023$.

Regression Analyses

Regressions are used to predict the likelihood of an outcome from a list of variables. In regressions, you can actually get a sense of how much of the variance in the dependent variable your predictor variables account for. Variance in the dependent variable equals how the dependent variable is different for different people. If there is no variance all people have the same score. If there is a lot of variance, different people have very different scores. How do we predict why different people have different scores?

In a recent study, researchers wanted to predict how good health formed (Fournier et al., 2017). Forty-eight students took part and attempted to practice a certain stretch (psoas-iliac stretch). The students practiced the stretch in the morning or the evening (one variable referred to as condition), consequently had different intentions to stretch (a second variable), and were men and women (a third variable). The researchers wanted to predict how these three variables predicted the students' levels of cortisol, a chemical that plays a role in habit formation. They conducted a regression analysis and their results table is reproduced in Table 2.3.

TABLE 2.3 ■ Regression Models Used to Determine the Mediating Role of Cortisol in the Effect of the Condition on the Time Taken to Form a Behavioral Habit

					95% CI	
Model and variable	**Coefficient**	***SE***	***t***	***p***	**LL**	**UL**
Model predicting cortisol: R^2 =.28, $F(1, 40)$ = 14.89, p =.0004						
Constant	1.627	.138	11.831	.000	1.349	1.905
Condition	.531	.138	3.859	.001	.252	.809
Model predicting $x_{.95}$ without inclusion of the mediator: R^2 =.20, $F(3, 38)$ = 2.34, p =.09						
Constant	288.495	172.121	1.671	.102	−59.950	636.941
Condition	−26.601	10.169	−2.616	.013	−47.187	−6.015
Intention	−30.528	35.453	−.861	.395	−102.299	41.244
Sex	−25.043	20.269	−1.236	.224	−66.075	15.990
Model predicting $x_{.95}$ with inclusion of the mediator: R^2 =.29, $F(4, 37)$ = 2.58, p =.053						
Constant	290.459	150.281	1.933	.061	−14.044	594.962
Cortisol	−22.413	10.779	−2.079	.045	−44.254	−.572
Condition	−14.218	11.213	−1.268	.213	−36.937	8.502
Intention	−23.578	31.366	−.752	.457	−87.133	39.976
Sex	−22.316	19.782	−1.128	.267	−62.398	17.767

Source: Fournier, M., d'Arripe-Longueville, F., Rovere, C., Easthope, C. S., Schwabe, L., El Methni, J., & Radel, R. (2017). Effects of circadian cortisol on the development of a health habit. Health Psychology, 36(11), 1059–1064. Reprinted with permission of APA.

Note: CI = confidence interval; LL = lower limit; ULCI = upper limit

Once more, you see a lot of numbers. Once more you pay the most attention to the *p*-values. There were three separate regression analyses represented by the three blocks of numbers. For our purposes, look at the first block of numbers. The variable that significantly predicted cortisol was the model with condition in it showing a direct effect of time of day on levels of cortisol. The other two models, the second and third blocks, show that neither sex nor intentions are significant variables (again look down the column beneath *p*).

Odds Ratios

One other statistical test that is relatively common in health psychology articles is the **logistic regression**. This analysis predicts the probability of the occurrence of an event. Articles will often report an **odds ratio**, which is the ratio of the odds of an event occurring in one group to the odds of it occurring in another group. Are men more likely to have a heart attack than women? (See Chapter 14 for the answer.) An odds ratio of 1 suggests the phenomenon (e.g., a heart attack) is equally likely in both groups. An odds ratio greater than 1 suggests the phenomenon is more likely to occur in the first group. Being comfortable with some commonly used statistics and analyses will make you a much better consumer of health psychology research and ultimately a better health psychologist. (See Field, 2018, for a good introduction to statistics and for more details on the terms discussed above.)

For a great example of how odds ratios look, here's an almost sad example. You know how many restaurants include so-called healthy options on their menu? Do you wonder if those options are presented in any different way than the regular items? Turnwald et al. (2017) took menus from 100 top-selling U.S. chain restaurants. They collected 262 healthy menu items with 5,873 words and 2,286 standard menu items with 38,343 words and measured if the type of words used in each type of menu was different. Table 2.4 reproduces a results table from their article and shows the use of odds ratios. As you see, restaurants described healthy items in less appealing ways. Look at the odds ratio column and you see the words more likely to occur in a standard menu were words such as *Exciting* OR = 3.26 and *Provocative* OR = 1.89. Words more likely to occur in a healthy menu were words such as *Simple* OR = 3.27 and, no surprise, *Nutritious* OR = 164.61. So healthy food choices are clearly portrayed differently and in a less appealing way. Perhaps this could be associated with those items being picked less.

TABLE 2.4 ■ Results of Healthy Menu Item Versus Standard Item Descriptions

Theme	Odds ratio [95% CI]	Log likelihood	Frequency in healthy menu (% of words)	Frequency in standard menu (% of words)
Words more likely to occur in standard menu				
Exciting	3.26 [1.73, 6.15]	19.26***	0.17	0.55
Fun and engaging	2.04 [1.56, 2.66]	33.11***	1.00	2.03
Traditional	1.96 [1.56, 2.47]	38.85***	1.35	2.61
American regional	1.96 [1.31, 2.92]	13.4***	0.44	0.86
Texture	1.95 [1.50, 2.54]	29.31***	1.02	1.98
Provocative	1.89 [.96, 3.73]	4.04*	0.15	0.29
Spicy hot	1.64 [1.12, 2.40]	7.29**	0.49	0.81
Artisan	1.63 [1.07, 2.48]	5.96*	0.41	0.67
Taste	1.52 [1.11, 2.08]	7.71**	0.75	1.13
Indulgent	1.37 [1.14, 1.65]	12.04***	2.21	3.01
No difference in healthy menu vs. standard menu				
Size	1.32 [.92,1.88]	2.46	0.58	0.76
Vague positive	1.27 [.77, 2.10]	.93	0.29	0.37
Choice	1.13 [.81, 1.57]	.54	0.68	0.77
Farm	1.20 [.87, 1.66]	1.15	0.73	0.61
Social	2.72 [.96, 7.72]	3.01	0.09	0.03
Words more likely to occur in healthy menu				
Foreign	1.27 [1.02, 1.58]	4.26*	1.62	1.28
Fresh	1.38 [1.09, 1.75]	6.39*	1.41	1.03
Simple	3.27 [1.68, 6.37]	10.25**	0.22	0.07
Macronutrients	8.76 [5.57, 13.77]	81.89***	0.75	0.09
Thinness	10.72 [7.22, 15.91]	134.28***	1.11	0.10
Deprivation	17.70 [8.56, 36.59]	68.68***	0.46	0.03
Nutritious	164.61 [40.04, 676.7]	185.49***	0.85	0.01

Source: Turnwald, B. P., Jurafsky, D., Conner, A., & Crum, A. J. (2017). Reading between the menu lines: Are restaurants' descriptions of "healthy" foods unappealing? *Health Psychology, 36*(11), 1034–1037. Reprinted with permission of APA.

Note: *p <.05; **p <.01; ***p <.001.

The statistic that is now overtaking the journals in health psychology is the **hazard ratio**. Both the odds ratio and the hazard ratio relate to relative risk. The probability of seeing a certain event in some group is called risk (Stare & Maucort-Boulch, 2016). The odds ratio is the odds of the probability of an event occurring in one group, divided by the probability of it not occurring. The hazard ratio is the comparison between the probability of events taking place in a treatment group compared to the probability of the events taking place in a control group. The hazard ratio essentially provides a statistical test of the efficacy of a treatment (Spruance et al., 2004). Almost every other health psychology journal article now seems to be reporting hazard ratios so be on the lookout for them.

Other important terms to watch for in the reporting of health psychological research are **relative risk**, the ratio of incidence or prevalence of a disease in an exposed group to the incidence or prevalence of the disease in an unexposed group, and **absolute risk**, a person's chance of developing a disease independent of any risk that other people may have.

Structural Equation Modeling

With an increase in technological sophistication, statistical tools allowed researchers to model multiple relationships simultaneously. A far cry from the zero-order correlation that maps the association of two variables, there are now analyses that can map out the relationships between an array of factors at the same time. One of the most popular is called structural equation modeling. As the name implies, you can draw a structure of variables and hypothesize how they are related. The statistical program then fits your model onto the data and generates an index to tell you how well your model fits the data. Not a great fit? Then, like trying on another piece of clothing, you can redraw your structure and try again.

Figure 2.3 shows you two structural equation models examining the relationships between trauma, resilience, and depressive symptoms on biological outcomes in African American smokers and nonsmokers (Berg et al., 2017). The solid black lines show the statistically significant associations and the gray lines show nonsignificant associations.

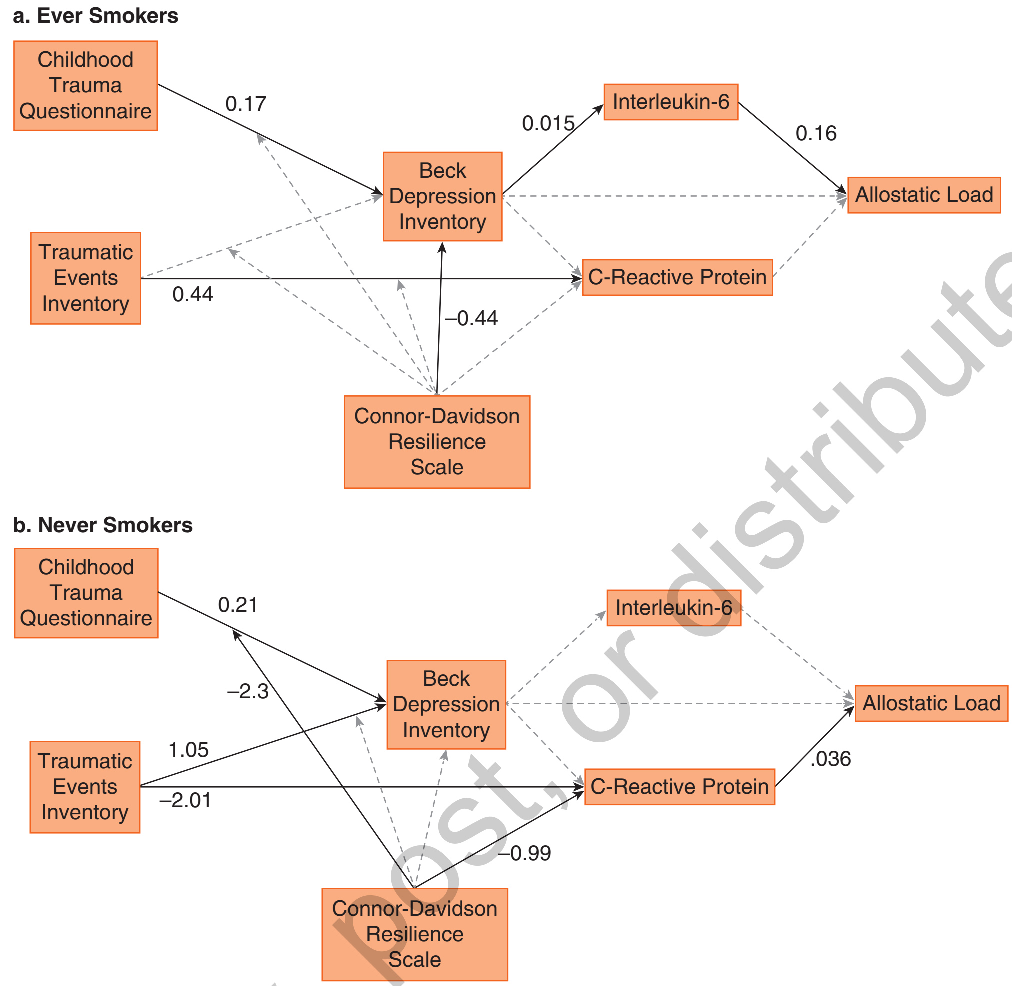
You now have all the key tips to make journal reading a much more pleasurable experience, not to mention a more educational one. Before diving into a broad discussion of how cultures vary in their approaches to health (Chapter 3), there is one more important discussion that will help you understand and interpret research better.

MODERATORS VERSUS MEDIATORS

Earlier in this chapter, I introduced you to the difference between zero-order correlations and partial correlations. The former looks at only two variables, the latter factors in or controls for others. I also introduced you to ANOVAs and MANOVAs. Even in those tests of group differences in variables you can also control for third variables. The resulting analyses are called ANCOVAs and MANCOVAs, where the "C" stands for covariance. Beyond just controlling for variables, health psychological research aims to test for the different ways that third variables can influence relationships between two other variables. If you have a large sample and the statistical chops, you can use structural equation modeling that maps out different relationships simultaneously. Before you do so, it is important to get comfortable with two types of roles variables can play. Welcome to the terms *mediation* and *moderation*. You will see many mediators and moderators in health psychology research.

Let's take the example of coping. There are many different ways coping can influence a health outcome. Although health psychologists originally focused on studying the direct relationships between stressors (e.g., public speaking) and outcomes (e.g., blood pressure), today researchers are paying more attention to underlying processes by which biopsychosocial factors influence health (Aldwin, 2019). Asking questions about specific effects (e.g., how, when, for whom, under what conditions, does public speaking lead to increased blood pressure?) requires moving beyond the examination of direct relationships to focus on additional factors that can explain how two variables are related. *Mediation* and *moderation* are two common examples of the type of processes now studied in detail across the field of health psychology (Talebi et al., 2016).

FIGURE 2.3 ■ Two Structural Equation Models



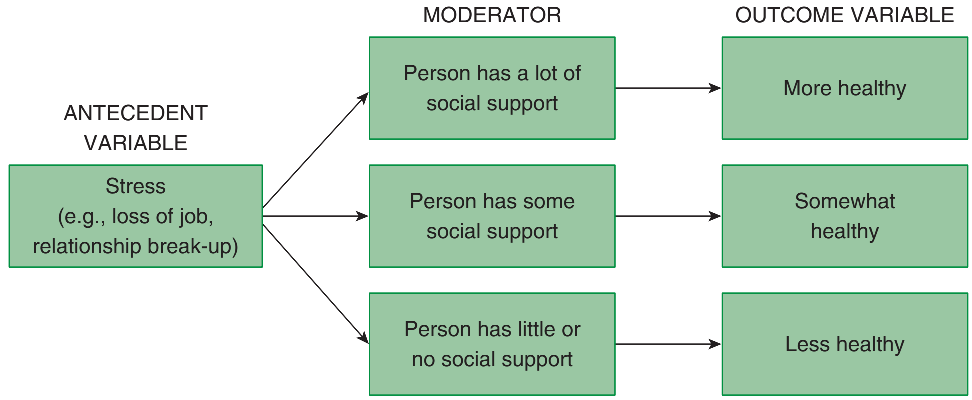
Source: Berg, C. J., Haardörfer, R., McBride, C. M., Kilaru, V., Ressler, K. J., Wingo, A. P., Saba, N. F., Payne, J. B., & Smith, A. (2017). Resilience and biomarkers of health risk in black smokers and nonsmokers. *Health Psychology, 36*(11), 1047–1058. Reprinted with permission of APA.

The context we are in can influence the things that first come to mind when we are asked to describe ourselves. If Manish were awakened from a nap and asked to describe himself, the order of things that would come up would be very different from those if he had an accident and were taken to the hospital. The context (the hospital) would bring different things to the level of consciousness. Being Hindu in a Western hospital may make those aspects of his self-concept more salient.

In life you will see that people who have a lot of a certain characteristic (or are high on that variable) tend to behave and react differently than people who have a little of that characteristic (or are low on that variable). The rich tend to be healthier than the poor. Older people tend to be more health conscious than younger people. People high in social support tend to cope better than people low in social support. In each of these cases the variable—income, age, and social support—are called moderators. A **moderator** is a variable that changes the *magnitude* (and sometimes the direction) of the relationship between an antecedent variable and an outcome variable (Aiken & West, 1991). This is easier to understand in a picture. Look at Figure 2.4.

In the example of social support, the number of stressors can be the antecedent variable and well-being is the outcome. A simple direct effect would be that people with more stressors are unhappier (a positive correlation). However, things are more complex than that. In any group of people, some individuals will have more social support than others. Let's measure social support and divide the people into a high-support and a low-support group. We would find that people with more support are

FIGURE 2.4 ■ Moderation

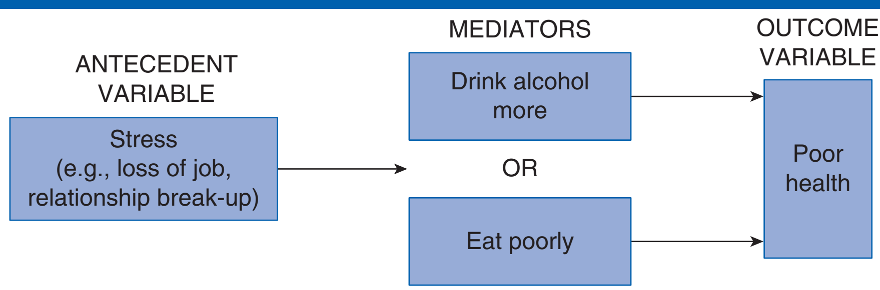


happier than people with less social support. Social support has moderated or buffered the relationship between stress and well-being. Such moderating effects of support are now well established and can be seen in a variety of life examples. For example, social support moderated the relationship between stress, mood, and alcohol use in a study of U.S. Navy personnel (Kelley et al., 2017). Being high or low on some factor often moderates how we react to stress.

Coping is often what you do when you are stressed. What you do can either help or hurt. The response to a stressor and the factors that follow a stressor influence what the outcome is going to be. These responses and factors between the stressor and the outcome are called mediators. A **mediator** is the intervening process (variable) through which an antecedent variable influences an outcome variable. Mediation can be described as a relationship where an independent variable changes a mediating variable, which then changes a dependent variable (MacKinnon, 2008). Coping behaviors in general and specific health behaviors are common mediators. Look at Figure 2.5. Instead of stress directly making you feel good or bad, it may influence your health behaviors (e.g., you drink alcohol or eat more) that *in turn* influence whether you feel good or bad. Here, health behaviors have mediated the relationship. Different cultures have different coping behaviors that can mediate the relationship between stress and well-being (Kuo et al., 2017). In one study of Chinese nurses, coping styles mediated the relationship between factors such as hope, optimism, resilience, and distress (Zhou et al., 2017).

A large body of literature in health psychology concerns interventions aimed at improving well-being by enhancing coping, based on the assumption that effective coping is a mediator (Coyne & Racioppo, 2000). There are a number of statistical procedures to test for mediation (Field, 2018). It is easy to see whether mediation is taking place by comparing the correlation between the antecedent and outcome variables before and after the potential mediator is entered into the statistical analysis. Stay with me here. If the variable you are studying is a mediator, the relationship between the antecedent and outcome variable significantly changes (gets lower) once the mediator is in the analysis. If you are stressed and you take a nap, you will probably wake up feeling better. If you are stressed

FIGURE 2.5 ■ Mediation



and you do not take a nap, you may feel worse. In this example, sleep is said to mediate the relationship between your stress level and how you feel. This chapter will discuss many other mediators and moderators.

Think about the different biopsychosocial variables we discussed so far and see whether you can tell the difference between moderators and mediators. Most variables that health psychologists study (e.g., coping styles and social support) can be both mediators and moderators. The role of the variable depends on the study (e.g., a cross-sectional study vs. a prospective study), the statistical analyses used to test the variable, the situation, or the variable under study. As a rule, mediators are changed by the stressor and correspondingly change the outcome. If more stress leads to you asking for more social support, which leads to you feeling better, social support is a mediator. If a longitudinal study shows that those with more stress exercise more and this makes them feel better, then exercise is a mediator. If a correlational study shows that the group of people who exercise more are less distressed than a similar group of stressed individuals who exercise less, then exercise is a moderator. In the first case (mediation), exercise follows the stressor, changing in level (e.g., increasing) and influencing the outcome. In the second case (moderation), we are looking at two separate groups of exercisers. The only variables that cannot be both mediators and moderators are those that cannot change as a function of the stressor or antecedent variable. Age, ethnicity, and race are examples of moderators that cannot be mediators (e.g., being more stressed cannot change your age even though some of us parents complain our kids' behavior age us faster).

If you catch yourself wondering what gets done when a variable such as social support can play the role of moderator and mediator, the good news is new statistical techniques allow us to test for such dual roles. In fact, sometimes (and to really increase complexity) you can have a variable moderate or mediate the relationship between other variables that are themselves moderators or mediators. This is referred to as moderated mediation, mediated moderation, or more commonly *conditional process analysis* (Hayes, 2022).

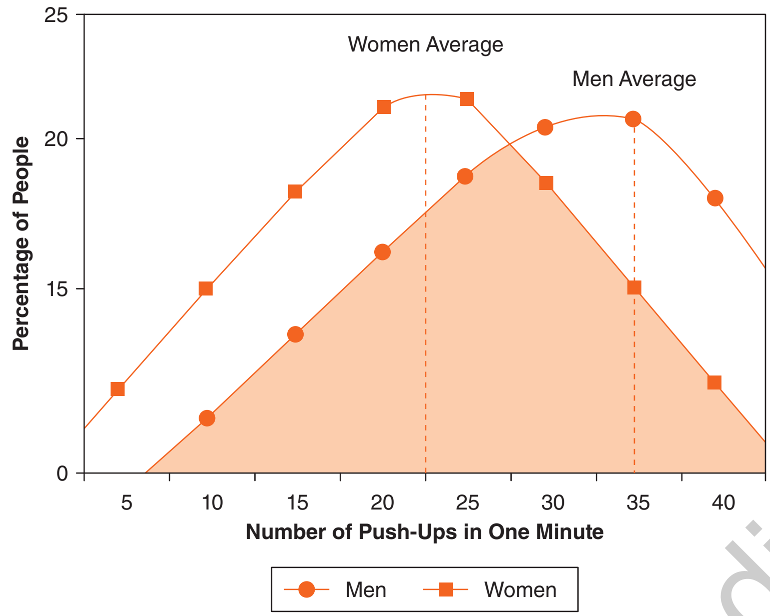
SOME FINAL CONSIDERATIONS

A key goal of this book is to present health psychology using a cultural approach. When we talk about culture, we often tend to emphasize cultural differences. To some extent, this is a natural human phenomenon. Even if people who are similar in age, ethnicity, and intelligence were to be randomly separated into two groups and forced to compete with each other, members of each group would tend to believe that they are better than those of the other group (e.g., the minimal subgroup paradigm, an important social psychological effect considered in later chapters). Even if we are not competing for resources, we still emphasize how we are different from other people.

There are two major problems here. First, this emphasis on differences often leads us to treat some groups better than others (factors such as prejudice are discussed later in this book). For example, we may be more likely to help people who look like us. We may be less likely to give information to someone who is not from a social group to which we belong. Second, whenever we deal with an individual from a culture with which we are not familiar, we are likely to use the key ways that they are different and generalize from that one person to the entire culture (this book later discusses the dangers of stereotyping as well). By focusing on major group differences, we often forget that differences exist within a group as often as between groups. Let's review an example.

Look at the two bell-shaped curves in Figure 2.6. The horizontal *x*-axis represents the number of push-ups a person can do, and the vertical *y*-axis represents the number of people who can do each number of push-ups. Now suppose we walk around town for a few days, and we ask every man and woman we see to get down on the pavement and do as many push-ups as they can in 1 minute. We continue this odd request until we have spoken to 100 men and 100 women. Each curve you see represents one of the two sexes. Therefore, the point of the curve for women above the number 10 means that of all the women we talked to (and who agreed to our strange request), 15 could do 10 push-ups. Now you will probably notice that the two curves are slightly set apart from each other. If we were to ask one of

FIGURE 2.6 ■ People Within Cultures Vary, Too



the most commonly asked questions in psychology, "Are there significant sex differences?" it is easy to see that the answer is yes. The average number of push-ups men can do is significantly higher than the average number of push-ups women can do. You can also see that there are more men who can do 30 or more push-ups than women, and more women than men who can only do 10 or fewer push-ups in 1 minute.

There are two critical things to notice about those two overlapping curves. First, even though there are men who can do more push-ups than any woman, and women who can do fewer push-ups than any man, notice how many men and women can do the same number of push-ups. The entire center portions of each curve overlap (the shaded part). At the heart of all this, we are all much more similar than we are dissimilar. Excluding unfortunate and unpredictable circumstances, we all have two eyes, two legs, a nose, and two ears. We all look pretty much the same. We all need to eat, drink, and sleep to live. So why then do we often look at either end of the curves or focus on group differences only? We do so because differences are more noticeable and provide a way to distinguish groups.

Even though we all need to eat, drink, and sleep to live, we vary in how we accomplish each of these activities, and how much food, drink, or sleep we need. These variations often make the difference between illness and health. This book will draw your attention to these variations. All humans have about 20,000 genes (compared with the fruit fly with 18,000 or the common earthworm at 12,000), but a variety of environmental and cultural factors can influence the kind of organisms those genes transcribe onto (Mukherjee, 2016). Humans share 99.8% of their genes, but that 0.2% difference is very important. In short, even though we should always remember that there are more similarities than differences, sometimes we can learn much from the differences.

There is something else to notice in Figure 2.6. Look at each curve by itself. Notice that there is a lot of diversity in push-up ability *even within each sex.* The average number of push-ups women can do provides a sense of general ability but notice the number of women at different stages of ability level. A lot of diversity is seen among women. This is a critical observation to hold on to as we discuss the many different topics in this book. No matter how many significant group differences we see, we must also remember that there are many differences within each culture as well. This basic understanding of the differences *within* versus *between* cultures applies to every culture we discuss. We can be talking about men and women as in the lighter push-up example above, or we can be comparing young and old, rich and poor, or Mexican American and African American. To make it easier to understand the different aspects of health psychology, this book will highlight how groups differ. Every time it does, keep these two overlapping curves at the back of your mind and always remember that these are average group differences only.

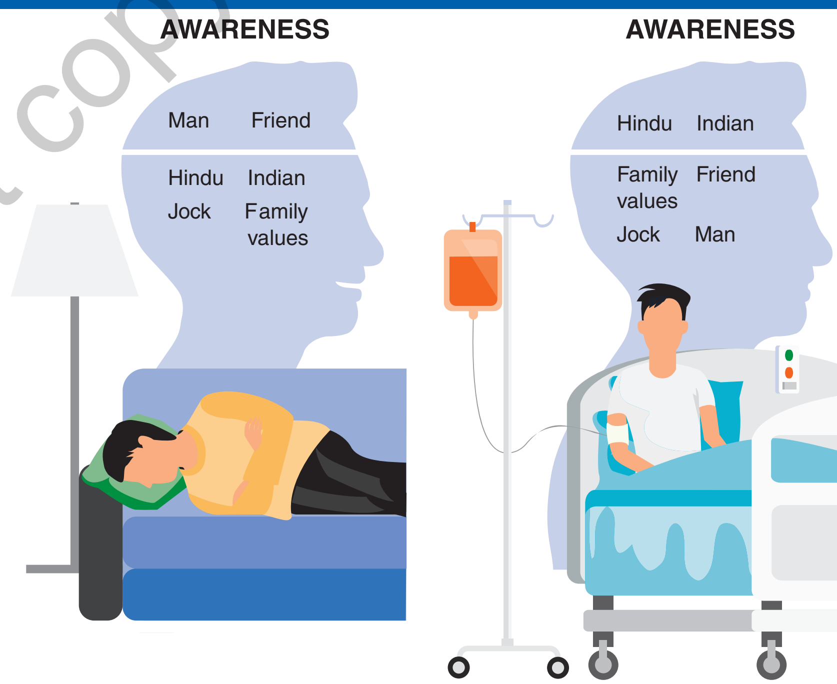
One final consideration. In Chapter 1, I had you answer the "Who am I?" test. If you (did what few readers do and actually) followed the question prompt, you would have a list of thoughts about who you were. The order in which the different descriptors came to your mind gives you a good idea of the aspects of yourself that are most important to you right now. It also alerts us to two critical factors to consider in this conversation of research design and measurement.

First, the order in which we use words to describe ourselves often depends on the **context** or the environment in which we are. If you are male and are answering the "Who am I?" question sitting in a room full of women, the answer, I am a man, is likely to be near the top of your list. Even if you did not answer with, I am American, your nationality probably would be one of the first descriptors that would come to mind if you were on a holiday abroad, say checking out the Tower of London surrounded by tourists from many different countries.

Even though the context can influence our ordering, it does not mean it changes the content of our self-views. This is where the level of analysis is important. This means that our views of ourselves reside at different levels of conscious awareness. Although you may think of yourself as a runner, this description may be far down on the list you generated and correspondingly we would have to go to a deeper level of analysis to uncover it. If we really want to get a good sense of a person and their culture, we have to remember that many different levels could be important and that the context in which we make our assessment can make a world of difference. Look at the example shown in Figure 2.7 and notice how the order of ways Manish describes himself varies depending on the context in which he is.

Having culture can offer a person many things. Think about what you may get from being part of a certain culture. Like someone in an army or someone on an athletic team (both cultures of their own), the culture in which you live influences ideas about what to do, what to wear, how to behave, and even how to feel. These prescriptions of how to be form the basis of the way the scientific literature defines culture.

FIGURE 2.7 ■ Levels of Analysis and Context



APPLICATION SHOWCASE: THE REPLICATION CRISIS IN PSYCHOLOGY

It has been a great few years for Wonder Woman. Movies featuring her have exceeded expectations at the box office, raking in millions. You may know of one way Wonder Woman is related to psychology. Her creator, William Moulton Marston, the same chap who invented the lie detector, was a psychologist. There is another way she is linked.

In 2010, Carney et al. published a paper that got people to sit up and take notice. Yes, both sit up and take notice. They demonstrated that our physical postures can influence a range of factors. In their study, they found that participants who assumed a pose conveying power, sitting back in a chair with feet on a desk and fingers laced behind their neck, actually felt better and released different hormones. One finding in particular caught the public's attention. The study claimed that holding a pose that expresses power and dominance, a power pose, before a meeting can influence psychological and physiological processes, as well as decision making. The classic analogy is standing like Wonder Woman with legs slightly apart and arms on hips.

The pose, and Amy Cuddy, went on to be featured on TED talks and in media around the world. Unfortunately, the findings of the 2010 study do not stand up to replication. In 2015, a group of researchers conducted a conceptual replication study using the same methodology as Carney et al. (2010) but using a much larger sample (200 vs. 42). They also strengthened their research design so the experimenter was blind, or unaware as to which condition the participants were in (Ranehill et al., 2015). The 2010 results did not replicate. In classic academic fashion, Carney et al. (2015) listed the many differences between the original 2010 paper and the 2015 failure to replicate that might have served as possible moderators. In 2017, 11 studies in two different journals failed to provide evidence of the original findings (Cesario et al., 2017). Along the way, Amy Cuddy was much maligned by her academic peers, a disturbing story well illustrated in a *New York Times Magazine* feature (Dominus, 2017) and with much of the intrigue and mayhem of a Hollywood box office hit.

The problem is not power poses. The problem is that a number of psychological findings fail to replicate. Another example involves the phenomenon known as *ego depletion*, the negative effect of performing a self-control task on performing another self-control task (Lurquin & Miyake, 2017). While a 2010 meta-analysis reported a moderate effect size (d = 0.62) for this phenomenon (Hagger et al., 2010), this finding too has shown replication failures (Lurquin et al., 2016), including a high-profile study involving 23 laboratories (Hagger et al., 2016).

The big exposé happened in 2015. Members of the Open Science Collaborative, more than 200 researchers led by Brian Nosek at Harvard, conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. They checked for reproducibility using p-values and effect sizes. Here is the shocker: Whereas 97% of original studies had significant results, only 36% of replications had significant results (Open Science Collaborative, 2015).

This issue, termed the replication crisis, also influences how we view health psychology research. It may be even more of an issue. For example, much public health research considers interventions that influence and are influenced by both individuals' health and the society around them. These complex systems necessitate explanation alongside statistical inference in light of the replication crisis (Grant & Hood, 2017).

Given the failures of replication, many people now look at the results of psychological research with caution. Some do not trust research. There are several statistical issues we should consider. One type of bias, known as **p-hacking,** occurs when researchers collect or select data or statistical analyses until nonsignificant results become significant. What the casual reader (or even the casual academic) fails to remember is that designing, collecting, analyzing, and reporting of psychological studies entail many arbitrary choices. Referred to as researcher degrees of freedom, these choices are problematic because they can influence the results. Wicherts et al. (2016) present an extensive list of 34 degrees of freedom that researchers have in formulating hypotheses, a great checklist.

There are other statistical issues as well. Researchers planning replication studies often use the original study sample effect size as the basis for sample size planning. However, this strategy ignores uncertainty and publication bias in estimated effect sizes, resulting in overly optimistic calculations. Anderson and Maxwell (2017) show that even if original studies reflect actual phenomena and are conducted in the absence of questionable research practices, popular approaches to designing replication studies may result in a low success rate, especially if the original study is underpowered.

I hope you are wondering if the different classic findings were false positives. Perhaps the replications showed that the emperor has no clothes and correctly indicates that there is truly no effect after all. One approach suggests failures to replicate may not be failures at all, but rather are the result of low statistical power in single replication studies. Statistical power, the likelihood that a study will detect an effect when there is an effect there to be detected, is another topic not discussed as much. Maxwell et al. (2015) provide examples of these power problems and suggest some solutions using Bayesian statistics and meta-analysis.

Measurement error adds noise to predictions and makes it more difficult to discover new phenomena. While our eyes always go to the *p*-value, it is important to remember statistical significance conveys very little information when measurements are noisy. This problem and related misunderstandings are key components in a feedback loop that perpetuates the replication crisis in science (Loken & Gelman, 2017).

A lesson to the wise. Recognize the pressures that may implicitly drive p-hacking, measurement issues, and failures to replicate. The growing emphasis on external funding as an expectation for faculty promotion may pose a large hazard for psychological science, including (a) incentives for engaging in questionable research practices, (b) a single-minded focus on programmatic research, (c) intellectual hyperspecialization, (d) disincentives for conducting direct replications, (e) stifling of creativity and intellectual risk taking, (f) researchers promising more than they can deliver, and (g) diminished time for thinking deeply (Lilienfeld, 2017).

Forewarned is forearmed. We live in an age when beyond not trusting social media shares of research studies, we need to focus on the statistical robustness of the studies as well.

CHAPTER REVIEW

Summary

- Health psychology uses the scientific method to design and plan research.
- There are many different types of research design and data collection methods. Research is primarily correlational or experimental in nature. Correlations are the assessment of association between variables. In experiments, researchers manipulate key variables (independent variables) to see the effects on others (dependent variables).
- The most common research design in health psychology is the randomized clinical trial. A special case of experiment, care is taken to select the target population and to ensure strong measurement.
- Common statistical tests include correlational analyses that assess the association between variables but can also control for third variables (partial correlations).
- ANOVAs and MANOVAs test for group differences while regression analyses predict variance in an outcome variable from a number of predictor variables.
- One of the most commonly seen analyses in health psychology today are odds ratio and hazard ratio analyses created using logistic regressions. Hazard ratios provide the likelihood or risk of a certain outcome in an intervention group as compared to a comparison group.
- Structural equation modeling allows one to simultaneously map the relationship between numerous variables and to ascertain the fit of a hypothesized set of relationships between variables with data collected.
- The context surrounding us can influence what we think about and how we see ourselves. Similarly, the deeper we analyze someone the more we learn about them.
- Although most research results, especially those discussing group differences, discuss averages, remember there can be a lot of difference across individuals within a group.

KEY TERMS

absolute risk
analyses of variance (ANOVAs)
context
correlation coefficient
correlational studies
cross-sectional
dependent variable
hazard ratio
incidence rates
independent variable
logistic regression
longitudinal
mediator
moderator
multivariate analyses of variance (MANOVAs)
odds ratio
p-hacking
partial correlation
placebo
prevalence rates
prospective
quasi-experimental designs
randomized, controlled, or clinical trials (RCTs)
regression analyses
relative risk
retrospective

ESSENTIAL READINGS

Benyamini, Y., Johnston, M., & Karademas, E. C. (Eds.). (2019). *Assessment in health psychology*. Hogrefe.

Luszczynska, A., Kruk, M., & Boberska, M. (2019). Measurement in health psychology research. In T. A. Revenson & R. A. R. Gurung (Eds.), *Handbook of health psychology* (3rd ed.). Routledge.

Randby, K. (2019). Major research designs in health psychology. In T. A. Revenson & R. A. R. Gurung (Eds.), *Handbook of health psychology* (3rd ed.). Routledge.