This chapter gives the reader an overview of major research designs utilized within the field of health psychology. Closely related to behavioral medicine (Freedland, 2017), health psychology research is also conducted within healthcare-related disciplines (e.g., public health, nursing, oncology, obstetrics, internal medicine). Research design in health psychology is a broad topic, particularly because of the diverse nature of the field and the various types of trainings health psychologists may have received. Rather than trying to present a representative view of all topics of study encompassed within health psychology, this chapter is organized around widely applicable questions that researchers should consider as they plan their research studies, including implications of manipulating versus measuring naturally occurring variables, as well as the ramifications of timing and frequency of measurement of variables. For example, if you were interested in understanding the impact of a cancer diagnosis on a person’s relationship functioning with their spouse, you would need to decide what theoretical constructs are important, how the constructs should be measured, whether one or both partners should be assessed or should receive an intervention, how long you should follow participants, and the number and timing of assessments. This chapter includes questions to consider when designing your study, a discussion of major types of research designs, and a discussion of types of measures.

Designing Your Study

What Is Your Program of Research?

To gain a full understanding of a topic within health psychology, research will likely involve qualitative and quantitative descriptive data, correlational data, experimental data, and dissemination research. Further, different types of data, such as self-report, observation, objective behavioral measures, big data (e.g., social media usage), medical records, and biological samples, reveal different aspects of phenomena. Additional chapters in this Handbook address measurement in health psychology research (Luszczynska, Kruk, & Boberska, 2019) and designing and conducting behavioral medicine interventions (Lovejoy & Fowler, 2019).

You should focus on the questions you are best equipped to answer. A single researcher does not need to perform all studies in a field of research. For example, a lab-based, experimental study could show that gain-framed messages increase physical activity (Gallagher & Updegraff, 2012), but it does not also need to show that physical activity leads to a reduction in cancer diagnoses. These
two questions reflect very different needs in terms of sample size, timing of assessments, length of follow-up, and expertise of the research team. Thus, they may best be answered utilizing different research designs.

Beyond an individual study, a researcher needs to consider how the study fits with his or her short- and long-term strategies for advancing a program of research. In other words, you should think about several studies that you would like to conduct that build upon each other. In this way, you gain clarity about the most important aspects of each individual study, which helps you identify what you most need to know before moving on to your next research question. This practice is widely recommended in academia and explicit in grant proposal writing as it ensures you have a good understanding of how single studies fit into the broader literature. Additionally, acquiring preliminary data for the next study within the current study is generally a good idea. For instance, testing a new measure or asking open-ended questions at the end of one study can produce helpful information to inform the design of a follow-up study.

**What Is Your Study’s Purpose?**

The guiding principle underlying most design decisions is your primary research question. The study’s purpose may be to gather descriptive data on a new problem or new measure. You may be interested in knowing how a construct changes over time, or you may want to show an association between two constructs. Alternately, you might want to demonstrate a causal effect between two or more variables. Trying to test too many research questions in a single study may result in a lack of ideal data to answer any of the questions.

**What Is the Target Population?**

Explicit in the process of defining a primary research question is defining your population of interest. In health psychology, much research concerns preventing illness or improving health; theories in both areas depend on the population of interest. A target population should match the study’s guiding theory and be defined by inclusion and exclusion criteria that may include disease severity, risk levels, and other relevant demographic factors. A population in which change can be observed on the primary outcome is another consideration. If a study is designed to prevent heart attacks, participants should be at high risk of experiencing heart attacks (e.g., Brancati et al., 2012). A psychosocial intervention to improve well-being may target a newly diagnosed population (e.g., Breitbart et al., 2015). Not only does it make conceptual sense that our research should focus on helping those most in need, choosing the appropriate population to sample improves your ability to find significant statistical effects.

Once you have determined your population of interest, you must carefully consider recruitment of your sample. Studies that aim to be representative of the population may utilize systematic sampling methods through institutions (e.g., schools to target adolescents) and sampling weights to correct for oversampling of minority groups (e.g., Harris et al., 2009). Beaver et al. (2017) demonstrated representative sampling by recruiting all Stage 1 endometrial cancer patients for a study testing the effect of two types of hospital follow-up on cancer patient’s mental health. Many studies may be unable to obtain a true random sample so convenience samples are often utilized. Convenience samples may be obtained online or through flyers to reach specific populations (e.g., designated smoking areas, oncology waiting rooms, pregnancy discussion boards). When convenience samples are used, the researcher must carefully consider the generalizability of the results. For example, research on community lifestyle interventions found different results depending on the recruitment method used (Estabrooks et al., 2017). In another example, research participation of caregivers of cancer survivors was found to relate to sexual minority status and relationship cohesion (Bazzi et al., 2016). Therefore,
a clear description of a study’s recruitment methods and sample characteristics is necessary for external validity and comparisons with other studies.

**Are There Important Subgroups in Your Population?**

In addition to having a clear definition of your population of interest, it is important to consider whether there are subgroups of people for whom your findings may differ. A stratified sampling technique may be utilized to allow for parameter estimates within each group. As an emphasis on having samples that are representative of demographics groups (e.g., gender, ethnicity, sexual orientation, age, socioeconomic status, marital status) continues to grow, you are likely to face decisions around inclusion/exclusion criteria and targeted recruitment to ensure certain groups are represented. The number of non-English speaking participants should continue to increase as measures are developed in other languages and technology improves to assist in translation and analysis of written text of multiple languages (e.g., Sung, Chang, Liu, Hsieh, & Chang, 2016).

If there is reason to believe that psychological processes will differ for certain groups, you may test whether an effect differs between groups, test effects separately within each group, or get one average effect across groups. Ideally, you should recruit enough people from all groups of interest so that you have statistical power to test moderation effects (Hayes, 2017). For example, you could test whether an intervention worked equally well for men and women. Interaction effects require larger sample sizes than bivariate effects and the more groups you want to test differences among, the larger your overall sample size will need to be.

Typically, sample homogeneity increases statistical power and clarifies generalizability. At the same time, it is critical for health psychologists to include diverse populations in research. The inclusion/exclusion criteria of a single study should therefore be guided by theory; moderation or subgroup analyses should be considered when possible.

**What Effect Do You Expect?**

Being able to frame your research questions in terms of specific, testable, hypotheses is a crucial early step in the design of your study. Most research questions involve testing effects concerning group differences or associations among two or more variables. You should specify your expected effect size based on other published studies, your own pilot work, or recommendations for clinically significant effects. Considering effect sizes in the design stage of your study will be incredibly helpful when considering your required sample size and when you publish your work, as it is widely recommended that the magnitudes of findings be reported in meaningful metrics (Pek & Flora, in press).

Large effects have advantages. Typically, they demonstrate more meaningful associations and require smaller sample sizes to detect statistical significance. A researcher may test a highly involved intervention to have a larger effect on an outcome than a more minimal intervention. Further, assessing a wider range of both predictor and outcome variables may provide evidence of a stronger association. At the same time, a small effect, obtained through limited resources, may be able to have a wide impact on population health. A review of public health interventions showed favorable cost-benefit ratios and return on investment estimates which is also important to consider (Masters, Anwar, Collins, Cookson, & Capewell, 2017).

**Is Your Study Sufficiently Powered to Detect Your Effect?**

Most research involves specifying a hypothesis and testing it using a null hypothesis significance test (NHST). NHSTs show that observed effects are greater than what would be expected due to chance if the null hypothesis were true. Conventional standards are that this chance, or probability, must
be less than 5% (i.e., p < .05) to reject a null hypothesis. The ability to detect effects if they exist at the population level is known as statistical power and increases as sample size increases. For more discussion of power, see Nuzzo (2016). Statistical programs (e.g., G-Power) help researchers estimate required sample size for detecting specified effects.

Concerns about researchers’ abilities to correctly utilize NHST have led to recommendations that range from supplementing tests with other information (e.g., effect sizes, graphic displays; Valentine, Aloe, & Lau, 2015) to abandoning p-values all together (Trafimow & Marks, 2015). An approach to analyzing data without the use of NHST is to use a Bayesian approach, which approximates the likelihood of an alternative hypothesis given a set of data (Buchinsky & Chadha, 2017).

What Is the Timing of the Effect?

Many studies in health psychology involve understanding how one construct affects another construct; however, the timing of these effects can vary widely. A causal variable may impact an outcome in a matter of seconds, weeks, or years. It is important to consider timing of a theoretical process so that you can best match interventions and measurement. Examining predictors of survival after a stroke necessitated following people for 10 years (Aparicio et al., 2017). For reducing stress with intervention message, smartphone technology can be used to deliver the intervention and assess the outcome instantaneously (Morrison et al., 2017).

What Is the Process of the Effect?

Health psychology theories are often more complex than simply X causes Y. Often, theories describe causal pathways through several constructs. For example, the Theory of Planned Behavior posits that attitudes, subjective norms, and perceived behavioral control cause intentions and, in turn, intentions cause performance of a behavior (Ajzen, 1985). Tests of theories of causal pathways (e.g., how does X lead to Y) require mediation analyses. Mediation analyses can be particularly useful when analyzing intervention effects in relation to an underlying theory and as a way to refine the intervention (MacKinnon, 2008). By estimating all hypothesized mediated effects in a single model, researchers can fully consider the impacts of the intervention on desired outcomes (Ranby et al., 2009, 2011).

Inherent in a mediation hypothesis is the concept that there is more than one causal association. The predictor (X) causes the mediator (M), which in turn, causes the outcome (Y). In addition to having a theoretical basis, research designs should be suited to address a mediation hypothesis. A weak approach to testing mediation hypotheses is with cross-sectional, correlational data because this is weak evidence of causality. Including additional time points, particularly ones theoretically chosen to match the hypothesized timing of change, can provide stronger evidence of a causal process (e.g., M preceding Y).

The best evidence of causal effects is obtained from experimental manipulation. Often, X is manipulated and M and Y are subsequently measured. In this way, strong causal evidence exists for the association between X and Y and for the association between X and M. The evidence of a causal link between M and Y, however, remains weak. It has been suggested that an X to Y to M, alternative model be tested to rule out reverse causality, however, Monte Carlo simulations have shown this is not a statistically sound approach (Lemmer & Gollwitzer, 2017). The strongest evidence that M causes Y comes from experimentally manipulating M. Direct manipulation of M may not be possible; however, a few examples exist (Bolger & Amarel, 2007; Word, Zanna, & Cooper, 1974).

Is Your Effect Likely to Replicate?

When we conduct research studies, we are interested in learning about broader populations than just our sample. In NHST we make explicit that decisions are inferences about a population based on
Major Research Designs in Health Psychology

A sample. When studies examine the same phenomenon within the same population, it is expected that the same results will typically be found. Failure to replicate may result from issues in the research design but may also be due to errors inherent in our statistical tests and publication bias. A primary goal of established methods of NHST is to minimize the chance of a Type 1 error. For a greater discussion of when this might be our goal as researchers, see Trafimow and Earp (2017). Although it might seem that minimizing the chance of a Type 1 error would help ameliorate the issue of replication, our field’s heavy reliance on p-values as determinates of statistical findings has had adverse effects. Namely, findings are typically only published if they are statistically significant, which has resulted in the overrepresentation of significant findings in a given area.

The phenomenon known as publication bias is problematic for the scientific field because professionals making policy or treatment decisions typically do not have access to studies with non-significant results. One recommendation for minimizing publication bias within clinical trials that is gaining widespread use, and now required by some journals, is to pre-register the trial (e.g., Moore, 2016). This ensures that there is a record of all trials performed in a given area, not only a record of trials with significant results. Further, through pre-registration, hypotheses are recorded a priori and not generated after significance tests are performed. For a broader discussion of replicability concerns, see Anderson and Maxwell, 2016.

Types of Designs

After thinking through these beginning questions, you should be starting to envision some of the main design features of your study. Which variables need to be manipulated? Which variables need measured more than once? For the next section of the chapter, relative strengths of different design features will be discussed. Unlike introductory textbooks that often define a study as “correlational” or “experimental”, you may benefit from thinking through a set of effects that you want to test when designing your study. In this way, you can consider characteristics of your study from the perspective of your variables. Which variables do you view as primary outcomes? Which variables do you want strong evidence of being causally related to your outcomes? Which variables do you need to statistically control for when explaining your outcome? How fast do you expect to see change on a particular outcome? If outcomes change at different rates, are both timing processes of equal importance to the study?

Experimental

The strongest evidence that two variables are causally related comes from an experimental design in which the researcher manipulates the level of one variable and subsequently measures the other. The essential features of experimental designs include randomly assigning participants to conditions and controlling extraneous factors so that the only difference among conditions is the level of the experimental factor. The following section highlights some innovative examples of experimental designs.

Participants may be randomly assigned to one condition at the start of a study, as is typical in experimental designs, but the condition they are assigned to may involve more than one manipulation. To examine the combined benefits of a diet and physical activity intervention, King and colleagues (2013) varied which of two interventions participants received as well as the timing of each. Adaptive interventions involve even more variability in the content, dose, or approach across persons and across time. For an introduction to adaptive interventions, see Collins (2018). A specific type of adaptive design, the Sequential Multiple Assignment Randomized Trial (SMART) design is useful for informing sequential, individualized approaches to managing health disorders. Two examples of SMART designs involve optimizing weight loss (Almirall, Nahum-Shani, Sherwood, &
Murphy, 2014) and optimizing pain management in cancer patients (Kelleher et al., 2017). An even more recent development in the area of adapted interventions are intensively adaptive interventions (IAIs), which utilize mobile technology to adjust quickly to the person’s individual state or social or environmental context (Riley, Serrano, Nilsen, & Atienza, 2015). Examples of these Just-In-Time adaptive interventions include adults with obesity being encouraged to take activity breaks from prolonged sitting at times when their smartphone detected sedentary behavior (Thomas & Bond, 2015) and adults who were quitting smoking being sent messages tailored to immediate lapse risk several times during the day (Hébert et al., 2018).

**Quasi-Experimental**

There are many constructs of interest to health psychologists that cannot be manipulated, either for practical or ethical reasons. Quasi-experimental designs may be used when researchers desire to test a causal hypothesis, but it is not possible to randomly assign participants to a condition. Given the lack of random assignment, quasi-experimental designs provide weaker causal evidence than experimental designs; however, within quasi-experimental designs, there is a range of design and statistical features that can strengthen evidence of causality. Some examples of quasi-experimental designs include examining the effectiveness of service delivery to pregnant smokers (Bennett, Jones, & Paranjothy, 2016) and examining whether changes in the built environment lead to reductions in childhood obesity (Goldsby et al., 2016).

**Correlational**

Correlational designs are useful for demonstrating an association among variables; correlations provide weak but necessary information for establishing a causal relationship. Correlations also aid in hypothesis generation and provide information on potential intervention targets for future research. It is easier to measure multiple potential predictor variables and an outcome variable than it is to manipulate each predictor variable. Predictors most strongly associated with an outcome variable could be considered as intervention targets if there is theoretical rationale. Correlational designs can be strengthened by the inclusion of multiple variables and multiple assessments.

**Longitudinal**

As compared to a cross-sectional design in which variables are only assessed at one time point, a longitudinal design indicates that variables were assessed from the same sample at more than one time point. Variables that are hypothesized to correlate with or predict the outcome variable can be measured once if they are believed to be stable constructs (e.g., demographics), or they can be measured many times if they are expected to change. The analyses used to examine longitudinal data may utilize constructs from the same time point or from different time points.

Measuring the same variable in the sample at multiple times has the statistical advantage of requiring relatively fewer people to detect the same size effect due to the ability to control for all individual differences over time (e.g., smaller error term, larger statistical test). Traditionally, longitudinal designs involve assessing participants weeks or months apart; although, some studies have followed people for many years. Longer studies can provide very useful information about health outcomes that take a long time to develop such as the effect of childhood bullying on mental health at age 50 (Evans-Lacko et al., 2017). Additionally, data collected for over 30 years showed perceived social support predicted later body mass index, controlling for their parent’s young adult body mass index (Serlachius et al., 2016).
Some effects, such as changing moods or knowledge, may happen very quickly, so measuring the outcome even minutes after the presentation of the independent variable may be appropriate. Many processes in health psychology occur on a time scale that is somewhere between instantaneous and months. Thus, it is important to align the timing of your assessment with the timing of the health process.

Many Time Points

For effects that are believed to occur rather quickly or multiple times a day, an intensive repeated measures design may be beneficial. These designs may also be referred to as ecological momentary assessments (EMAs), which highlights the fact that they are occurring in natural settings. For example, one study assessed participants hourly for seven days to assess the association between impulse control and snacking behavior (Powell, McMinn, & Allan, 2017). Another study examined variations in health behaviors, affect, and academic performance among first year university students using 61 brief surveys; the first 28 surveys were sent every six days and the last 33 were sent daily during end-of-year examinations (Flueckiger, 2017). This design highlights matching assessments to meaningful times when change is hypothesized to occur.

Studies involving multiple processes that occur at different times may consider a measurement burst design. This design combines a frequent assessment (e.g., hourly, daily) with an extended period of following participants (e.g., months, years). In order to reduce participant burden while obtaining necessary data to examine both kinds of research questions, the frequent assessments occur for shorter periods of time, spaced throughout the study. For instance, Piasecki, Hedeker, Dierker, and Mermelstein (2016) utilized this design to examine smoking behavior and mood among adolescents. The simultaneous association of smoking behavior and mood was assessed up to five times a day for one week. These one-week assessments periods were repeated four times, spaced over 24 months during high school, a time when substance use patterns change. In another study, emotional stress reactivity was examined using daily assessments for 30 days, repeated annually for four years (Howland, Armeli, Feinn, & Tennen, 2017).

N of 1

An N of 1 design is an approach typically utilized to understand the effect of an intervention in which a single person receives a random allocation of the experimental and control conditions. This design can be an effective way to demonstrate causality without a comparable control condition. The strongest evidence of causality among these types of designs is an ABA withdrawal experimental design that assesses an individual before the intervention, during the intervention, and after the intervention. For an example of an N of 1 design within osteoarthritis, see O’Brien, Philpott-Morgan, and Dixon (2016). With more emphasis being placed on individualized medicine, the use of these types of designs may be on the rise within health psychology, particularly when they are used in conjunction with more traditional experimental approaches (Lillie et al., 2011).

Types of Measures

Another important aspect of research design is utilizing measures that most closely reflect constructs of interest. Hypothesis formation should involve operationalizing constructs, and chosen measures should assess the construct with high validity and reliability. Keep in mind that teams of researchers are becoming increasingly common in health psychology, so if your research requires the use of an unfamiliar measure that takes a lot of training or specialized equipment (e.g., fMRI, biomarkers,
structured interviews), you may benefit from finding a collaborator. There are many kinds of measures in health psychology. Four broad types will be discussed: self-report, objective behaviors, big data, and biological data.

**Self-Report Measures**

Traditionally, health psychologists have widely employed self-report measures due to their relative low cost and ease of analysis. Further, the person’s own perception is often the theoretical construct of interest (e.g., their own well-being, support available to them in their social network). Self-reported data may be collected through paper or online survey and in-person or telephone interviews. Many health psychology scales are collected on closed-ended, Likert-type scales which allow for quantitative analyses. Open-ended questions are particularly important when researchers are not able to anticipate responses. Commonly, individuals participate in research alone, but some questions may benefit from focus groups with several participants or from dyadic participation.

Self-report measures have limitations, the severity of which depends on the research question. Self-report measures are prone to non-conscious (e.g., memory failures) and conscious (e.g., self-presentation, demand characteristics) biases. If the person’s perception is not the construct of interest, they may be poor reporters. Within the social support literature, distinguishing between who is reporting the support (the recipient or the provider) (Bolger, Zuckerman, & Kessler, 2000) and the kind of support (available or received) has important implications for theoretical effects (Uchino, 2009). It might be necessary to collect data from other people in the person’s life. For example, a partner report might be useful for assessing interpersonal interactions, and a parent, teacher, or friend report might be useful for assessing adolescent substance use.

**Objective Behavioral Measures**

Research on health behaviors that impact prevention or treatment of disease is increasingly emphasizing objective rather than self-reported measures (e.g., Celis-Morales et al., 2012). Many studies have incorporated tracking applications (apps) on smartphones. Further, the increase in wearable devices that automatically track behavior is particularly exciting as they eliminate participant biases in perception and self-presentation. These devices are used frequently in physical activity research (Coughlin & Stewart, 2016) as well as cigarette smoking (Sazonov, Lopez-Meyer, & Tiffany, 2013), alcohol use (Kim et al., 2016), and sleep (Brunet, Dagenais, Therrien, Gartenberg, & Forest, 2017). A comparison of devices shows the performance of the device depends on the primary behavior of interest (physical activity vs. sleep; Rosenberger, Buman, Haskell, McConnell, & Carstensen, 2016). Medication adherence can also be objectively monitored with electronic pill bottles (Sutton et al., 2014).

**Naturally Occurring “Big Data”**

A fourth kind of data that is increasingly being utilized in health psychology is referred to as “big data”. Similar to archival data, big data refers to data that is originally collected for non-research purposes. As data is stored electronically, the amount of data recorded and potentially available to researchers is immense. Psychologists have analyzed language data from social media (Kern, 2016), investigated negative emotions related to traumatic events on Twitter (Jones, Wojcik, Sweeting, & Silver, 2016), and used hashtags to study declarative memory (Stanley & Byrne, 2016).

Physiological indicators obtained from smart clothing and health-related applications provide new opportunities (Chen, Ma, Song, Lai, & Hu, 2016). Large-scale data sets allow for unique opportunities to find naturally occurring patterns and associations (Chen & Wojcik, 2016; Kosinski, 2016).
These data also have limitations; they may not be able to be linked to other important variables, limiting the research questions that can be tested.

**Biological Measures**

Finally, biological measures reflect an important aspect of health. They may be obtained through medical records or primary analysis of blood, saliva, or tissue samples. Heart rate from medical records was compared to post-traumatic stress symptoms following injury (Marsac, 2017). Biomarkers (interleukin-6 [IL-6], C-reactive protein [CRP], allostatic load) obtained from whole blood have been compared between smokers and non-smokers (Berg, 2017). A protein marker in saliva (C-Reactive Protein), an index of inflammation, has been examined in conjunction with interactions between parents and adolescents (Nelson, 2017). Telomeres, caps at the end of each strand of DNA in our cells, protect our chromosomes and shorten as we age. Stress and poor health behaviors shorten telomeres (Shammas, 2011) as do negative emotions from the family environment (Brody, 2017) and discrimination (Lee, 2017).

**Conclusions**

Health psychologists consider effects at the biological level, the individual level, the interpersonal level, the cultural level, and the policy level. Knowing the strengths of your training and infrastructure, in addition to the current literature in your area, can help you frame your research questions. Increasingly common in health psychology is conducting research within an interdisciplinary team where people with differing expertise can design various aspects of the larger study. Taking your time to plan all aspects of the research design and analytical plan will benefit your study, your program of research, and ultimately the field.

**References**


Krista W. Ranby


