9

Multivariate Correlational Research

CORRELATIONAL STUDIES CAN PROVIDE interesting new information in their own right. The opening headlines provide examples. It might interest us to read that children who are praised too much are also self-centered or narcissistic. We might be surprised to learn that watching sex on TV shows is linked to teen pregnancy. Often, however, a correlational result is an early step in establishing a causal relationship between two variables. Psychological scientists (among many others) want to know about causes and effects, not just correlations, because they may suggest treatments. If praise is linked to narcissism, we might wonder whether or not the praise *makes* kids narcissistic. If it does, parents might change how they express approval. When reading that sexual content on TV is linked to teenage pregnancy, we may wonder whether watching sexual material *causes* behavior that leads to pregnancy. If it does, then pediatricians, teachers, or advocacy groups could argue for restricting teens' exposure to certain kinds of TV shows. However, if the relationships are not causal, such interventions would not work.

Because correlation is not causation, what are the options? Researchers have developed some techniques that enable them to test for cause. The best of these is experimentation: Instead of measuring both variables, researchers manipulate one variable and measure the other. (Experimental designs are covered in Chapters 10–12.) Even without setting up an experiment, however, researchers can use some advanced correlational techniques to get a bit closer to

LEARNING OBJECTIVES

A year from now, you should still be able to:

1. State why simple bivariate correlations are not sufficient for establishing causation.

2. Explain how longitudinal correlational designs can establish temporal precedence.

3. Explain how multiple-regression analyses can rule out some (but not all) third variables.

4. Describe the value of pattern and parsimony, in which a variety of research results support a single, parsimonious causal theory.

5. Explain the function of a mediating variable.

making a causal claim. This chapter outlines three such techniques: longitudinal designs, which allow researchers to establish temporal precedence in their data; multiple-regression analyses, which help researchers rule out certain third-variable explanations; and the "pattern and parsimony" approach, in which the results of a variety of correlational studies all support a single, causal theory. In the three techniques, as in all correlational studies, the variables are measured—that is, none are manipulated.

REVIEWING THE THREE CAUSAL CRITERIA

The bivariate examples in Chapter 8 involved only two measured variables. In contrast, longitudinal designs, multiple-regression designs, and the pattern and parsimony approach are **multivariate designs**, which involve more than two measured variables. While these techniques are not perfect solutions to the causality conundrum, they are extremely useful and widely used tools, especially when experiments are impossible to run.

Remember that the three criteria for establishing causation are covariance, temporal precedence, and internal validity. We might apply these criteria to correlational research on the association between parental praise and narcissism.

In the research you'll read about in this chapter, narcissism is studied as a personality trait in which people feel superior to others, believe they deserve special treatment, and respond strongly when others put them down. Parental overpraise, the other variable discussed in this example, occurs when parents tell kids they are exceptional or more special than other children. It's important to note that childhood narcissism is different from high self-esteem (a trait that is considered healthy). Similarly, overpraising is different from parents expressing warmth and love for their children.

Let's examine the three criteria:

1. *Is there covariance?* At least one study did find covariance (Otway & Vignoles, 2006). Adults who were narcissistic remembered their parents praising them for almost everything they did. The correlation was around $r = .20$.
2. *Is there temporal precedence?* A correlational study like Otway and Vignoles's does not establish temporal precedence because both variables were measured at the same time. In a single session, adults rated their narcissism and also reflected on their parents' past behavior. Therefore, their current self-views could have colored their recall of the past. It's not clear which variable came first in time.
3. *Is there internal validity?* The association between parental praise and child narcissism might be explained by a third variable. Perhaps parents praise boys more than girls, and boys are also more likely to have narcissistic traits. Or perhaps parents who are themselves narcissistic simply overpraise their children and, independently, their narcissism is mimicked by their kids.

CHECK YOUR UNDERSTANDING

1. Why can't a simple bivariate correlational study meet all three criteria for establishing causation?

1. See p. 242.

ESTABLISHING TEMPORAL PRECEDENCE WITH LONGITUDINAL DESIGNS

A **longitudinal design** can provide evidence for temporal precedence by measuring the same variables in the same people at several points in time. Longitudinal research is used in developmental psychology to study changes in a trait or an ability as a person grows older. In addition, this type of design can be adapted to test causal claims.

Researchers conducted such a study on a sample of 565 children and their mothers and fathers living in the Netherlands (Brummelman et al., 2015). The parents and children were contacted four times, every 6 months. Each time, the children completed questionnaires in school, responding to items about narcissism (e.g., "Kids like me deserve something extra"). Parents also completed questionnaires about overpraising their children, which was referred to in the study as overvaluation (e.g., "My child is more special than other children").

This study was longitudinal because the researchers measured the *same* variables in the *same* group of people across time—every 6 months. It is also a multivariate correlational study because eight variables were considered: child narcissism at Times 1, 2, 3, and 4, and parental overvaluation at Times 1, 2, 3, and 4.

Interpreting Results from Longitudinal Designs

Because there are more than two variables involved, a multivariate design gives several individual correlations, referred to as cross-sectional correlations, autocorrelations, and cross-lag correlations. The Brummelman researchers conducted their analyses on mothers' and fathers' overvaluation separately, in order to investigate the causal paths for each parent separately. We present the results for mothers here, but the results were similar for fathers.

CROSS-SECTIONAL CORRELATIONS

The first set of correlations are **cross-sectional correlations**; they test to see whether two variables, measured at the same point in time, are correlated. For example, the study reports that the correlation between mothers' overvaluation

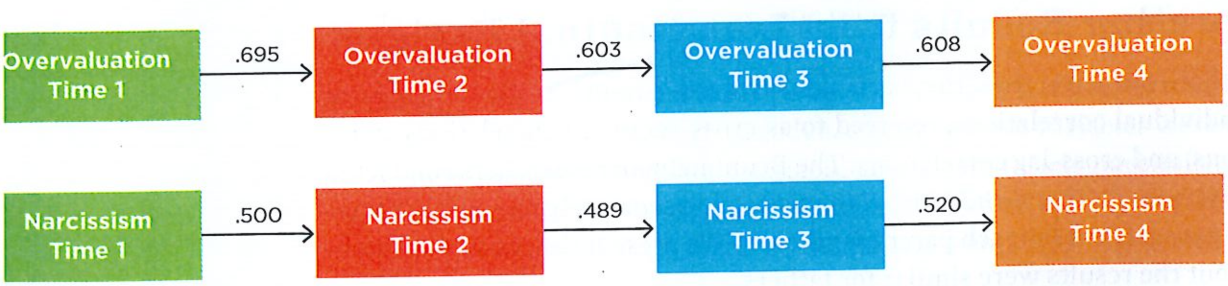
URE 9.1
ss-sectional correlations.
k at the correlations of the variables when measured at the same time. Within each time period, the mothers' rvaluation is weakly associated with child narcissism. Notice that the arrows point in both directions because ese cross-sectional correlations, the two variables were measured at the same time, so we don't know which e first. The figure shows zero-order (bivariate) correlations. (Source: Adapted from Brummelman et al., 2015.)

at Time 4 and children's narcissism at Time 4 was $r = .099$. This correlation is consistent with the hypothesis. However, because both variables in a cross-sectional correlation were measured at the same time, this result alone cannot establish temporal precedence. Either one of these variables might have led to changes in the other. **Figure 9.1** depicts how this study was designed and shows all of the cross-sectional correlations.

AUTOCORRELATIONS

The next step was to evaluate the correlation of each variable with itself across time. For example, the Brummelman team asked whether mothers' overvaluation at Time 1 was associated with mothers' overvaluation at Times 2, 3, and 4; they also asked whether children's narcissism at Time 1 was associated with their scores at Times 2, 3, and 4. Such correlations are sometimes called **autocorrelations** because they determine the correlation of one variable with itself, measured on two different occasions. The results in **Figure 9.2** suggest that both overvaluation and narcissism are fairly consistent over time.

» Autocorrelations are the e as test-retest reliability rrelations (see Chapter 5).



URE 9.2
tocorrelation.
a longitudinal study, researchers also investigate the autocorrelations. These results indicate that both iables seem to be relatively stable over time. Notice that the arrows point in only one direction, because are sure that the Time 1 measurements came before the Time 2 measurements. (Source: Adapted from ummelman et al., 2015.)

CROSS-LAG CORRELATIONS

So far, so good. However, cross-sectional correlations and autocorrelations are generally not the primary interest. Researchers are usually most interested in **cross-lag correlations**, which show whether the earlier measure of one variable is associated with the later measure of the other variable. Cross-lag correlations thus address the directionality problem and help establish temporal precedence.

In the Brummelman study, the cross-lag correlations show how strongly mothers' overvaluation at Time 1 is correlated with child narcissism later on, compared with how strongly child narcissism at Time 1 is correlated with mothers' overvaluation later on. By inspecting the cross-lag correlations in a longitudinal design, we can investigate how one variable correlates with another one (that's the "cross" part of its name) over time (that's the "lag" part). Cross-lag correlations establish temporal precedence. In Brummelman's results, only one set of the cross-lag correlations was greater than zero; the other set was not (**Figure 9.3**).

Statistically significant correlations. Notice that Brummelman's team reported the point estimates for the correlations but did not report the 95% CIs for each. Instead, they used the shorthand of statistical significance, which is related to the 95% CI. Perhaps the correlation of $r = .071$ in Figure 9.3 (the association between Time 1 overvaluation and Time 2 narcissism) had a 95% CI of [.01, .15], which does not include zero. As you learned in Chapter 8, when a 95% CI for a correlation does not include zero, researchers can also say that the correlation is "statistically significant." In contrast, perhaps the correlation from Time 1 narcissism to Time 2 overvaluation had a 95% CI of [–.09, .07]. Because this CI does include zero, we can also say that the correlation is "not significant," abbreviated "n.s." Brummelman's team used this shorthand instead of reporting the 95% CIs for each correlation.

Taken together, the cross-lag correlations mean that mothers who overvalued their children at one time had children who were higher in narcissism 6 months later. In contrast, children who were higher in narcissism at a particular time did

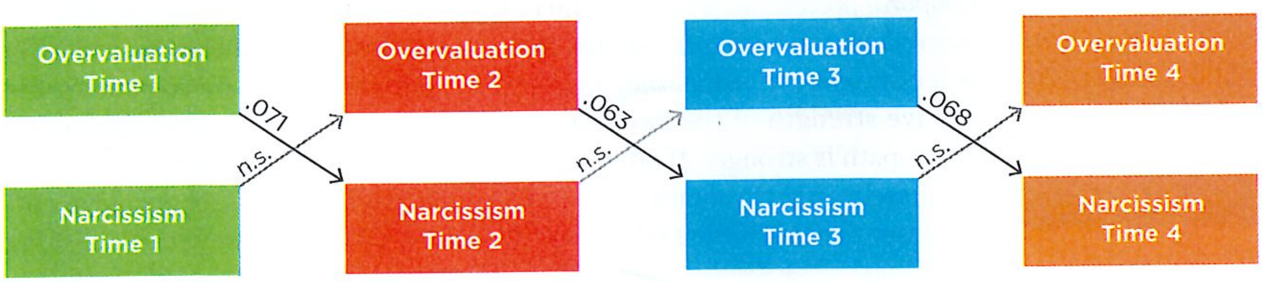


FIGURE 9.3
Results of a cross-lag study.
The cross-lag correlations in this study are consistent with the conclusion that parental overpraise comes before narcissism because overpraise in early time periods significantly predicts later narcissism, but narcissism in earlier time periods was not significantly (n.s.) related to later overpraise. The arrows point in only one direction because in each case the method makes clear which variable came first in time; Time 1 always comes before Time 2, and so on. Values shown are associated with mothers' overpraise. (Source: Adapted from Brummelman et al., 2015.)

not have mothers who overvalued them 6 months later. Because the "overvaluation-to-narcissism" correlations are significant and the "narcissism-to-overvaluation" correlations are not, the results suggest the overvaluation, not the narcissism, came first.

Three Possible Patterns from a Cross-Lag Study. The results of the cross-lag correlations in the Brummelman study could have followed one of three patterns. The study did show that parental overpraise (overvaluation) at earlier times was correlated with child narcissism at the later times. Such a pattern was consistent with the argument that overpraise leads to increases in narcissism over time. However, the study could have shown the opposite result—that narcissism at earlier times was correlated with overpraise later. Such a pattern would have indicated that the childhood narcissistic tendency came first, leading parents to change their type of praise later.

Finally, the study could have shown that *both* correlations were different from zero—that overpraise at Time 1 predicted narcissism at Time 2 *and* that narcissism at Time 1 predicted overpraise at Time 2. If that had been the result, it would mean excessive praise and narcissistic tendencies are mutually reinforcing. In other words, there is a cycle in which overpraise leads to narcissism, which leads parents to overpraise, and so on.

Longitudinal Studies and the Three Criteria for Causation

Longitudinal designs can provide some evidence for a causal relationship by means of the three criteria for causation:

1. *Covariance.* Statistical relationships in longitudinal designs help establish covariance. When two variables are correlated and their 95% CIs do not contain zero (as in the cross-lag correlations in Figure 9.3), there is covariance.
2. *Temporal precedence.* A longitudinal design can help researchers make inferences about temporal precedence. Because each variable is measured at clearly different points in time, they know which one came first. By comparing the relative strength of the two cross-lag correlations, the researchers can see which path is stronger. If only one of them is statistically significant (as in the Brummelman overvaluation and narcissism study), the researchers move a little closer to determining which variable comes first and closer to establishing causation.
3. *Internal validity.* When conducted simply—by measuring only the two key variables—longitudinal studies may not help rule out third variables. For example, the Brummelman results presented in Figure 9.3 cannot clearly rule out the possible third variable of socioeconomic status. It's possible that parents in higher income brackets overpraise their children, and that children in upper-income families are also more likely to think they're better than other kids.

However, researchers can sometimes design their studies or conduct subsequent analyses in ways that address some third variables. For example, in the Brummelman study, one possible third variable is gender. What if boys show higher levels of narcissism than girls, and what if parents of boys are also more likely to overpraise them? Gender might be associated with both variables. Participant gender does not threaten internal validity here, however, because Brummelman and his colleagues report that the pattern was the same when boys and girls were examined separately. Thus, gender is a potential third variable, but by studying the longitudinal patterns of boys and girls separately, the Brummelman team was able to rule it out.

Why Not Just Do an Experiment?

Why would Brummelman and his team go to the trouble of tracking children every 6 months for 2 years? Why didn't they just do an experiment? After all, conducting experiments is the only certain way to confirm or disconfirm causal claims. The problem is that in some cases people cannot be randomly assigned to a causal variable of interest. For example, we cannot manipulate personality traits, such as narcissism in children. Similarly, while parents might be able to learn new ways to praise their children, they can't easily be assigned to daily parenting styles, so it's hard to manipulate this variable.

In addition, it could be unethical to assign some people, especially children, to a condition in which they receive a certain type of praise, especially over a long time period, particularly if we suspect that one type of praise might make children narcissistic. Similarly, if researchers suspect that smoking causes lung cancer or sexual content on TV causes pregnancy, it would be unethical (and difficult) to ask study participants to smoke cigarettes or watch certain TV shows for several years. When an experiment is not practical or ethical, a longitudinal correlational design is a good option.

Nevertheless, researchers who investigate how children react to different types of praise have not relied solely on correlational data. They have developed ethical experiments to study such reactions, at least over a short time period (Brummelman et al., 2016; Mueller & Dweck, 1998). By randomly assigning children to receive praise for who they are (e.g., "You are so smart") versus praise for how hard they worked (e.g., "You must have worked hard at these problems"), researchers have produced some solid evidence that children really do change their behavior and attitudes in response to adult praise (**Figure 9.4**). Because it is ethically questionable to expose children to potentially harmful feedback, such studies had to pass strict ethical review and approval before they were conducted (see Chapter 4). In addition, the exposure time was short (only one instance of praise per study, and no instances of criticism). It would be much more challenging to do an ethical experimental study of the effects of long-term

FIGURE 9.4
Praising children.
Correlational and experimental studies suggest that when adults praise children's learning strategies and efforts (compared with praising the type of person they are), kids respond favorably and continue to work hard.

exposure to potentially maladaptive praise at home. That makes longitudinal correlational designs an attractive alternative.

CHECK YOUR UNDERSTANDING

1. Why is a longitudinal design considered a multivariate design?
2. What are the three kinds of correlations obtained from a longitudinal design? What does each correlation represent?
3. Describe which patterns of temporal precedence are indicated by different cross-lag correlational results.

1. See p. 243. 2. See pp. 243–246. 3. See pp. 245–246.

RULING OUT THIRD VARIABLES WITH MULTIPLE-REGRESSION ANALYSES

> Groundbreaking research suggests that pregnancy rates are much higher among teens who watch a lot of TV with sexual dialogue and behavior than among those who have tamer viewing tastes. (CBSNews, 2008)

This news item, referring to a study on TV content and teenage pregnancy, reports a simple association between the amount of sexual content teens watch on TV and their likelihood of becoming pregnant (Chandra et al., 2008). But is there a causal link? Does sexual TV content *cause* pregnancy? Apparently there is covariance: According to the published study, teens who watched more sexual material on TV were more likely to get pregnant. What about temporal precedence? Did the TV watching come before the pregnancy? According to the report, this study did establish temporal precedence because the researchers first asked teens to report the types of TV shows they like to watch and followed up 3 years later with the same teens to find out if they had experienced a pregnancy.

What about internal validity? Third variables could explain the association. Perhaps one is age: Older teenagers might watch more mature TV programs, and they're also more likely to be sexually active. Or perhaps parenting is a third variable: Stricter parents might monitor their teens' TV use and also put tighter limits on their behavior.

How do we know whether one of these variables—or some other one—is the true explanation for the association? This study used a statistical technique called **multiple regression** (or *multivariate regression*), which can help rule out some third variables, thereby addressing some internal validity concerns.

Measuring More Than Two Variables

In the sexual TV content and pregnancy study, the researchers investigated a sample of 1,461 teenagers on the two key variables (Chandra et al., 2008). To measure the amount of sexual TV content viewed, they had participants report how often they watched 23 programs popular with teens. Then coders watched 14 episodes of each show, counting how many scenes involved sex, including passionate kissing, sexually explicit talk, or intercourse. To assess pregnancy rates 3 years later, they asked girls, "Have you ever been pregnant?" and asked boys, "Have you ever gotten a girl pregnant?" The two variables were positively correlated: Watching higher amounts of sex on TV was associated with a higher risk of pregnancy (**Figure 9.5**).

If the researchers had stopped there and measured only these two variables, they would have conducted a bivariate correlational study. However, they also measured several other variables, including the total amount of time teenage participants spent watching any kind of TV, their age, their academic grades, and whether they lived with both parents. By measuring all these variables instead of just two (with the goal of testing the interrelationships among them all), they conducted a multivariate correlational study.

USING STATISTICS TO CONTROL FOR THIRD VARIABLES

By using a multivariate design, researchers can evaluate whether a relationship between two key variables still holds when they **control for** another variable. To introduce what "controlling for" means, let's focus on one potential third variable: age. Perhaps sexual content and pregnancy are correlated only because older teens are both more likely to watch more mature shows and more likely to be sexually active. If this is the case, all three variables are correlated with one another: Viewing sex on TV and getting pregnant are correlated, as we already determined, but

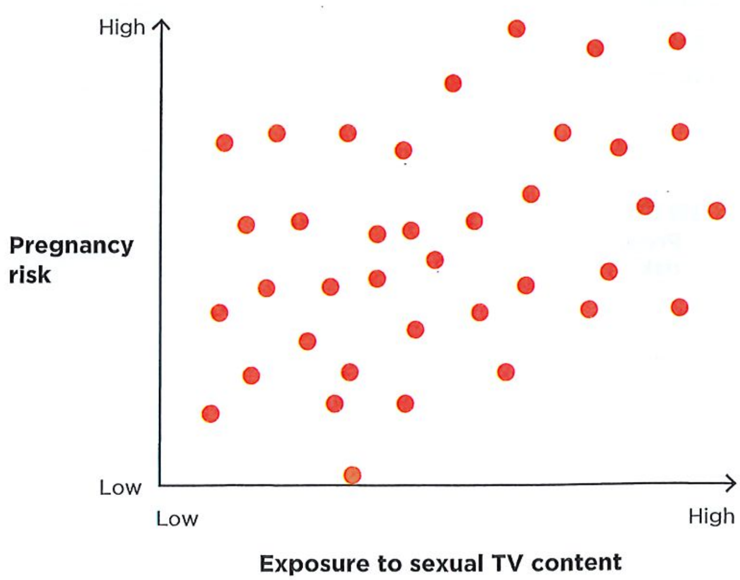


FIGURE 9.5
Correlating sexual TV content with pregnancy risk.
Higher rates of sexual content on TV go with higher risk of pregnancy, and lower rates of sexual content go with lower risk of pregnancy. (Data are fabricated for illustration purposes.)

sex on TV and age are also correlated with each other, and age and pregnancy are correlated, too. The researchers want to know whether age, as a third variable correlated with both the original variables, can account for the relationship between sexual TV content and pregnancy rates. To answer the question, they see what happens when they control for age.

You'll learn more about multiple-regression computations in a full-semester statistics course. This book will focus on a conceptual understanding of what these analyses mean. The most statistically accurate way to describe the phrase "control for age" is to talk about proportions of variability. Researchers are asking whether, after they take the relationship between age and pregnancy into account, there is still a portion of variability in pregnancy that is attributable to watching sexy TV. But this is extremely abstract language. As an analogy, you might compare this variability to the overall movement (the variance) of your wiggling, happy dog when you return home. You can ask, "What portion of the variability in my dog's overall movement is attributable to his tail moving? To his shoulders moving? To his back legs moving?" You can ask, "Will the dog still be moving when he greets me, even if I were to hold his tail constant—hold it still?"

Another way to understand "controlling for" is to recognize that testing a third variable with multiple regression means identifying subgroups. We can think of it like this: We start by looking only at the oldest age group (say, 20-year-olds) and see whether viewing sexual TV content and pregnancy are still correlated. Then we move to next oldest group (age 18), then the youngest group (age 16). We ask whether the bivariate relationship still holds at all ages.

There are a couple of possible outcomes from such a subgroup analysis, and one is shown in the scatterplot in **Figure 9.6**. Here, the overall association is positive—the more sexual TV programs teens watch, the higher the chance of getting pregnant. In addition, the oldest teens (the 20 symbols) are, overall, higher on sexual

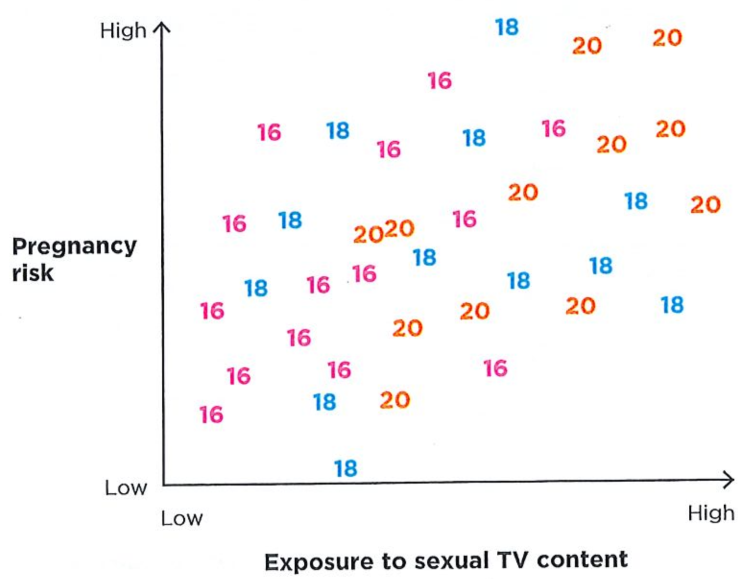


FIGURE 9.6

The association between sexual TV content and pregnancy remains positive, even controlling for age.

The overall relationship shown here is positive, and this holds even within the three subgroups: age 20, age 18, and age 16. (Data are fabricated for illustration purposes.)

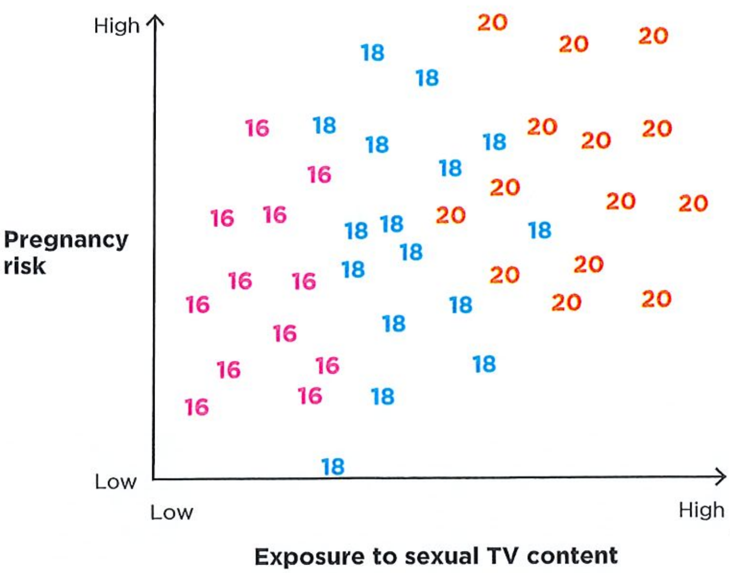


FIGURE 9.7
The association between sexual TV content and pregnancy goes away, controlling for age.
The overall association shown here is positive, but if we separately consider the subgroups of age 20, age 18, or age 16, there is no relationship between the two variables. (Data are fabricated for illustration purposes.)

TV content and higher in chance of pregnancy. The youngest teens (16 symbols) are, overall, lower on sexual TV content and lower in chance of pregnancy. If we look *only* at the 20-year-olds, or *only* at the 16-year-olds, however, the scatterplots still show the key relationship between sexy TV and pregnancy: It remains positive even within these age subgroups. Therefore, the relationship is still there, even when we hold age constant.

In contrast, the second possible outcome is shown in **Figure 9.7**. Here, the overall relationship is still positive, just as before—the more sexual content teens watch on TV, the higher the chance of pregnancy. In addition, just as before, the 20-year-olds watch more sexy TV and are more likely to become pregnant. However, this time, when we look *only* at the age 20 subgroup or *only* the age 16 subgroup, the key relationship between sexy TV and pregnancy is absent. The scatterplots *within* the age subgroups do not show the relationship anymore. Therefore, the association between watching sexual TV content and getting pregnant goes away when we control for age. In this case, age was, indeed, the third variable that was responsible for the relationship.

Regression Results Indicate Whether a Third Variable Affects the Relationship

Which one of the two scatterplots, Figure 9.6 or Figure 9.7, best describes the relationship between sexual content on TV and pregnancy? The statistical technique of multiple regression can tell us. When researchers use regression, they are testing whether some key relationship holds true even when a suspected third variable is statistically controlled for.

As a consumer of information, you'll probably work with the end result of this process, when you encounter regression results in tables in empirical journal

articles. Suppose you're reading an article and you come across a table showing what the regression results would look like for the sexy TV/pregnancy example. What do the numbers mean? What steps did the researchers follow to come up with them?

CRITERION VARIABLES AND PREDICTOR VARIABLES

When researchers use multiple regression, they are studying three or more variables. The first step is to choose the variable they are most interested in understanding or predicting; this is known as the **criterion variable**, or *dependent variable*. The Chandra team were primarily interested in predicting pregnancy, so they chose that as their criterion variable. The criterion (dependent) variable is usually specified in either the top row or the title of a regression table, such as **Table 9.1**.

The rest of the variables measured in a regression analysis are called **predictor variables**, or *independent variables*. In the sexy TV/pregnancy study, the predictor variables are the amount of sexual content teenagers reported viewing on TV and the age of each teen. In Table 9.1, the two predictor variables are listed below the criterion variable.

USING BETA TO TEST FOR THIRD VARIABLES

The point of the multiple-regression results in Table 9.1 is to see whether the relationship between exposure to sex on TV and pregnancy might be explained by a third variable—age. Does the association remain, even within each age group (as in Figure 9.6)? Or does the relationship between sexy TV and pregnancy go away within each age group (as in Figure 9.7)? The betas in Table 9.1 help answer this central question.

BLE 9.1

ultiple-Regression Results from a Study :edicting Pregnancy from Sexual Content on V and Age

ITERION EPENDENT) RIABLE: EGNANCY RISK	BETA	95% CI FOR BETA	STATISTICAL SIGNIFICANCE
edictor ıdependent) riables:			
posure to sex on TV	0.25	[.14, .36]	*
ıe	0.33	[.20, .46]	*

: Data are fabricated, based on imagined results if the researchers had used two predictor variables.
.05, meaning the result is statistically significant and the 95% CI does not de zero.

Beta Basics. In a regression table like Table 9.1, there is often a column labeled beta (or β, or even standardized beta). There will be one beta value for each predictor variable. Beta is similar to r, but it reveals more than r does. A positive beta, like a positive r, indicates a positive relationship between that predictor variable and the criterion variable, when the other predictor variables are statistically controlled for. A negative beta, like a negative r, indicates a negative relationship between two variables (when the other predictors are controlled for). A beta that is zero, or nearly zero, represents no relationship (when the other predictors are controlled for). Therefore, betas are similar to correlations in that they denote the direction and strength of a

relationship. The higher beta is, the stronger the relationship is between that predictor variable and the criterion variable. The smaller beta is, the weaker the relationship.

«
For more on confidence intervals of beta, see Statistics Review: Inferential Statistics, p. 513.

Within a single regression table, we can usually compare predictor variables that show larger betas to predictor variables with smaller betas—the larger the beta, the stronger the relationship. For example, in Table 9.1 we can say that the beta for the age predictor appears a bit stronger than the beta for the exposure to sex on TV predictor. Keep in mind, however, that it is not appropriate to compare the strengths of betas from one regression table to the strengths of betas from another one. The reason is that betas change, depending on what other predictor variables are being used—being controlled for—in the regression (Rohrer, 2018; Westfall & Yarkoni, 2016).

Sometimes a regression table will include the symbol b instead of beta. The coefficient b represents an unstandardized coefficient. A b is similar to beta in that the sign of b—positive or negative—denotes a positive or negative association (when the other predictors are controlled for). But unlike two betas, we cannot compare two b values within the same table to each other. The reason is that b values are computed from the original measurements of the predictor variables (such as dollars, centimeters, or inches), whereas betas are computed from predictor variables that have been changed to standardized units. A predictor variable that shows a large b may not actually denote a stronger relationship to the criterion variable than a predictor variable with a smaller b.

Interpreting Beta. In Table 9.1, notice that the predictor variable "exposure to sex on TV" has a beta of 0.25. This positive beta, like a positive r, means higher levels of sex on TV go with higher pregnancy risk (and lower levels of sex on TV go with lower pregnancy risk), *even when we statistically control for the other predictor on this table—age*. In other words, even when we hold age constant statistically, the relationship between exposure to TV sex and pregnancy is still there. This result is consistent with the relationship depicted in Figure 9.6, not the one in Figure 9.7.

The other beta in Table 9.1, the one associated with the age predictor variable, is also positive. This beta means that older age is associated with higher pregnancy rates, *when exposure to sex on TV is controlled for*. In other words, when we hold exposure to sex on TV constant, age predicts pregnancy, too. In sum, the beta that is associated with a predictor variable represents the relationship between that predictor variable and the criterion variable, when the other predictor variables in the table are controlled for.

95% CIs and Statistical Significance of Beta. The regression tables in empirical journal articles, especially those published in recent years, have a column labeled 95% CI, which presents the confidence interval for each beta. Regression tables published longer ago may only have a column labeled sig or p, or may have an asterisked footnote giving a p value for each beta. Recall that a p value of .05 complements the .95 from a 95% CI. Specifically, when the p value is less

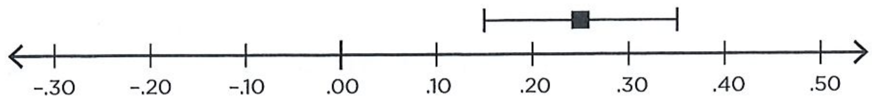


FIGURE 9.8

Confidence intervals and statistical significance of beta.

When the 95% CI does not include zero, we can say the beta is "statistically significant."

than .05, you can infer that the 95% CI for that beta does not contain zero and is therefore considered statistically significant. When p is greater than .05, the beta is considered not significant (n.s.), and you can infer that its 95% CI *does* contain zero (**Figure 9.8**).

Table 9.1 contains both pieces of information (even though they mean the same thing). Each of the betas reported there has a 95% CI that does not include zero, and both are noted as $p < .05$, or statistically significant. Both columns share similar information, but the 95% CI is more informative because it communicates the precision with which beta is estimated (narrower CIs are more precise). **Table 9.2** gives several appropriate ways to explain what the beta for the TV variable means.

What If Beta Is Close to Zero? To answer this question, we'll use an example from a different line of research: family meals and child academic achievement. When these two variables are studied as a bivariate relationship, researchers find that children in families that eat many meals together (dinners and breakfasts) tend to be more academically successful, compared with kids in families that eat only a few meals together.

Once again, this simple bivariate relationship is not enough to show causation. In many studies, family meal habits and academic success are measured at the same time, so there is a temporal precedence problem: Did family meals come first and reinforce key academic skills, leading to higher achievement? Or did high academic success come first, perhaps making it more pleasant for parents to have

TABLE 9.2

Describing the Beta of 0.25 in Table 9.1

EACH OF THESE SENTENCES IS AN APPROPRIATE DESCRIPTION OF THE RELATIONSHIP:
• The relationship between exposure to sex on TV and pregnancy is positive (high levels of sex on TV are associated with higher levels of pregnancy risk), even when age is controlled for.
• The 95% CI for the relationship between exposure to sex on TV and pregnancy does not contain zero, suggesting that this relationship is positive, controlling for age.
• The relationship between exposure to sex on TV and pregnancy is positive (high levels of sex on TV are associated with higher pregnancy risk) and is not attributable to the third variable of age because it holds even when age is held constant.

TABLE 9.3

Multiple-Regression Results from a Study Predicting Academic Success from Frequency of Family Meals and Parental Involvement

CRITERION (DEPENDENT) VARIABLE: ACADEMIC SUCCESS	BETA	95% CI FOR BETA	STATISTICAL SIGNIFICANCE
Predictor (independent) variables:			
Frequency of family meals	−0.01	[-0.06, 0.03]	n.s.
Parental involvement	0.09	[0.06, 0.12]	*

Note: Data are fabricated but reflect actual research. The study controlled for not only parental involvement but also income, family structure, school quality, birth weight, school type, and many other possible third variables. When controlling for all these in a sample of more than 20,000 children, the researchers found that the beta for frequency of family meals was not significant.

*$p < .05$, meaning the result is statistically significant and the 95% CI does not include zero.

Source: Adapted from Miller et al., 2012.

meals with their kids? In addition, there are third variables that present an internal validity concern. For instance, more involved parents might arrange more family meals, and more involved parents might also have higher-achieving children.

A multiple-regression analysis could hold parental involvement constant and see if family meal frequency is still associated with academic success. In one such study, the researchers found that when parental involvement was held constant (along with other variables), family meal frequency was no longer a strong predictor of school success (Miller et al., 2012). This pattern of results means that the only reason family meals were correlated with academic success was because of the third-variable problem of parental involvement (**Table 9.3**).

In other words, although frequency of family meals and academic success are related in their bivariate relationship, that relationship goes away when potential third variables, such as parental involvement, are controlled for. When you hold parental involvement constant, there is no longer a relationship between frequency of family meals and academic success (**Table 9.4**).

TABLE 9.4

Describing the Beta of −0.01 in Table 9.3

EACH OF THESE SENTENCES IS AN APPROPRIATE DESCRIPTION OF THE RELATIONSHIP:

- When controlling for parental involvement, the relationship between family meal frequency and child academic success has a 95% CI that contains zero (is not significant).
- The relationship between family meal frequency and child academic success can likely be explained by the third variable of parental involvement.
- The relationship between family meal frequency and child academic success goes away when parental involvement is controlled for.

Adding More Predictors to a Regression

Up to now, when considering the relationship between sexual TV content and pregnancy, we've focused on only one potential internal validity problem—age. But remember there are many other possible third variables. What about participation in school activities? What about living with one versus two parents? In fact, the Chandra team measured each of those third variables and even added a few more, such as parental education, ethnicity, and having a history of problem behaviors (Chandra et al., 2008). **Table 9.5** shows every variable tested, as well as the multiple-regression results for all the other variables. Because the Chandra team did not report 95% CIs, only the statistical significance is listed here.

Even when there are many more predictor variables in the table, beta still means the same thing. The beta for the exposure to sex on TV is positive: High levels of sex on TV are associated with higher pregnancy rate, when the researchers controlled for age, total TV exposure, lower grades, parent education, educational aspirations, and so on, down to intention to have children before age 22. Even after controlling for all variables listed in Table 9.5, the researchers found that more exposure to sex on TV predicts a higher chance of pregnancy.

TABLE 9.5

Multiple-Regression Results from a Study Predicting Pregnancy from Exposure to Sex on TV and Other Variables

CRITERION (DEPENDENT) VARIABLE: PREGNANCY RISK	BETA	SIG
Predictor (independent) variables:		
Exposure to sex on TV	0.44	*
Total television exposure	-0.42	*
Age	0.28	*
Lower grades	0.21	n.s.
Parent education	0.00	n.s.
Educational aspirations (highest level of school you plan to finish)	-0.14	n.s.
Being Hispanic (vs. other ethnicities)	0.86	n.s.
Being Black (vs. other ethnicities)	1.20	*
Being female	1.20	*
Living in a 2-parent household	-1.50	*
History of deviant or problem behavior (e.g., skipping school, stealing, cheating on a test)	0.43	*
Intention to have children before age 22	0.61	n.s.

$p < .05$, meaning that the 95% CI for this beta does not contain zero.
Source: Adapted from Chandra et al., 2008, Table 2.

Adding several predictors to a regression analysis can help answer two kinds of questions. First, it helps control for several third variables at once. In the Chandra study, even after all other variables were controlled for, exposure to sex on TV still predicted pregnancy. A result like that gets the researchers a bit closer to making a causal claim because the relationship between the suspected cause (sexy TV) and the suspected effect (pregnancy) does not appear to be attributable to any of the other variables that were measured.

Second, by looking at the betas for all the other predictor variables, we can get a sense of which other factors predict chance of pregnancy. One strong predictor is gender, which, as you can see, has a beta of 1.20, even when the other variables are controlled for. This result means girls are more likely to report becoming pregnant than boys are to report getting a girl pregnant. (Even though it takes two to cause a pregnancy, presumably boys are sometimes unaware of getting a girl pregnant, whereas a girl is more certain.) We also notice that teens with a history of deviant behavior also have a higher risk of pregnancy, controlling for exposure to sex on TV, age, grades, and

the other variables in the table. In fact, the predictive power of history of deviant behavior is about the same magnitude as that of exposure to sex on TV. Even though the authors of this study were most interested in describing the potential risk of viewing sexual content on TV, they were also able to evaluate which other variables are important in predicting pregnancy. (Recall, however, that when a table presents *b* values, or unstandardized coefficients, it is not appropriate to compare their *relative* strength. We can only do so with beta, and even then, remember that betas change depending on what other predictor variables are used.)