CHAPTER 22

DYADIC DATA ANALYSIS

Richard Gonzalez and Dale Griffin

The study of interdependence contributes to a rich understanding of social life. Does a husband's depression influence his wife's depression? Does his depression influence her marital satisfaction? What is the similarity of the husband and wife's depression within a couple and does that similarity predict other variables such as marital satisfaction? What predicts the degree of similarity in depression between husband and wife? Research questions such as these involve data that span two individuals, so we say the data are *interdependent*. Such research questions frequently include multiple variables and sometimes involve longitudinal data. What makes these research questions psychologically interesting is that they focus on interpersonal processes. Of course, if one wants to study interpersonal processes, then it would be useful to collect data and use analytic procedures that permit the assessment and testing of interpersonal processes.

The analysis of interdependent data presents special issues because the covariance across individuals needs to be addressed in the analyses. Failure to account for these interpersonal correlations can introduce bias into an analysis but, more important, consideration of these interpersonal correlations allows one to assess interesting interpersonal psychological processes. The violation of independence is the ugly pebble that can be transformed into the pearl of interdependence. In this chapter, we illustrate a few analytic techniques that go beyond "fixing" data for independence violations to providing rich models that permit the researcher to assess psychological processes of interdependence. Interdependence is not treated as a nuisance that needs to be corrected but rather as one of the key psychological parameters to model. We view this chapter as introductory, focusing on the special case of dyads, and review a few of the analysis techniques that are currently available.

The analysis of dyadic data can become detailed in that there are many issues that need to be addressed in the analysis, such as whether dyad members are exchangeable or distinguishable, whether dyadic data are cross-sectional or longitudinal, whether one wants to frame the analysis as a multilevel model or a structural equation model (SEM), whether one takes a latent variable approach to individual and dyadic variance or whether one's theoretical model focuses more on direct relations across people and across variables on observed data, whether data are normally distributed or categorical, and so on. As with any data analysis exercise, these different design features and priorities yield many combinations that are too numerous to review in a short introductory chapter. We refer the reader to a comprehensive treatment by Kenny, Kashy and Cook (2006), which covers many of these different combinations, and the edited volume by Card, Selig, and Little (2008).

Our goal for this chapter is to highlight some simple concepts so researchers can develop an intuition and an appreciation for the procedures that are possible when one includes both individuals from a dyad in the same analyses. We discuss some of our favorite issues surrounding dyadic analyses and review a handful of the available data analytic

DOI: 10.1037/13621-022
APA Handbook of Research Methods in Psychology: Vol. 3. Data Analysis and Research Publication, H. Cooper (Editor-in-Chief)

techniques, taking a more intuitive and basic approach rather than presenting a general framework in which the designs are special cases.

STUDY OF INTERDEPENDENCE VERSUS VIOLATION OF INDEPENDENCE

We initially became interested in dyadic data analysis because as social psychologists we wanted to study dyadic processes but recognized that the analytic techniques available at the time were too limiting. It seemed strange to us that researchers who studied dating couples, for example, collected data from only one individual in the couple. This practice occurred partly because it was a way to bypass the violation of independence. One automatically has independent data when only one member of each couple is represented in the data set. Can a researcher who is interested in studying dating couples make claims about the couple from data from only one person? Rather than go down the road of "one-hand clapping" philosophical-type arguments, we decided instead to work on analytic procedures that allowed the study of the couple.

A study that assesses only one dating partner from each dyad gathers information about the relationship, as indexed by that single subject. One can learn about the husband's marital satisfaction or the wife's satisfaction but not how each dyad member influences the other's marital satisfaction. A study that assesses both members of the couple can examine how one person influences another, can examine the similarity or dissimilarity of the couple members, can examine whether the degree of similarity predicts some other variable, and so on. The potential for testing richer psychological questions increases dramatically when the researcher collects data from both individuals. One can now study the dyad in a deeper way.

The analysis of dyadic data has gone through an identity crisis over the past two decades. In the early days, methodologists warned about the dangers of failing to properly account for dependency in data when observations are made on both members of a dyad. The basic idea was that when members of the same dyad are analyzed together, their data are dependent on each other in much the same way that two observations in a two-time-point repeated measures design are related to each other. We understand well what happens to statistical inference if we ignore temporal dependence in a repeated measures analysis, and the early literature on dyadic data analysis focused on the analogous effects of ignoring nonindependence in dyadic research (e.g., Kenny, 1990; Kenny & Judd, 1986). This early literature made use of the intraclass correlation (ICC) as a tool to model the effects of violating independence. The ICC can be used to show how classic analysis of variance (ANOVA) and regression designs go awry in the presence of nonindependent data, such as how parameters are biased or when *p* values become either too liberal or too conservative.

Concern about the violation of independence seems to be a small piece of the larger puzzle. If one wants to study couples, then interdependence comes with the territory. Interdependence is part of the phenomenon, not a statistical flaw in one's data. The problem of violating independence is not with the data or the research question, but it is merely a symptom that one is using an inappropriate analytic technique for the research question. A researcher of dyadic processes should not have to adjust their interdependent data, or their interdependent research question, to conform to a statistical model of independence. Instead, the statistical tools should allow the researcher to model interdependence directly in much the same way that a researcher has tools for analyzing data that are temporally dependent. The tools should provide ways to assess and measure interdependence, and should provide ways to test theories of interdependence. In short, statistical techniques should facilitate the goals of the researcher, not provide roadblocks that get in the way of testing theory and understanding process.

Fortunately, the field of dyadic data analysis has moved away from primarily being concerned about the bias introduced through violations of the independence assumption. The current focus is on developing methods that facilitate the testing of research questions about interdependence. We next turn to a general introduction to types of association in data, which will provide a grounding for understanding dyadic data analysis.

TEMPORAL, INTERPERSONAL, AND MULTIVARIATE CORRELATIONS

There are three common types of associations that occur in psychological data. One type is temporal association. For example, data from the same participant are collected multiple times. We do not treat those observations as independent because we want to capture the temporal association, and in many cases, the temporal association is the key focal point. We do not merely use repeated measures or longitudinal designs because they provide more statistical power; we use repeated measurements because they provide unique information about change processes. A second type of association is due to interrelations among multiple variables. Such associations across variables are what multiple regression and SEMs assess. What is the relation between marital satisfaction and depression? We use multivariate techniques to study the association between variables. The third type of association is interpersonal association as seen in dyadic designs. Observations may have interpersonal associations because they come from the same members of a social unit, such as the two members of a dyad, or students in the same classroom. We use dyadic techniques because we are interested in studying interpersonal processes.

These three kinds of associations between observations are not mutually exclusive, with one, two, or all three possibly occurring in the same study. For example, if there is a single dependent variable for each member of the dyad, then the observations are correlated by virtue of the interpersonal relations. If those same observations are also repeated for each dyad member, now there are associations both interpersonally and temporally in the data. If the dyad members are each measured once but on different variables, then both multivariate and interpersonal associations exist in the data. If both dyad members are observed repeatedly over time across several variables, then the observations exhibit associations temporally, interpersonally, and across variables. Analyses should account for all types of associations present in the data set.

Fortunately, there is a single basic idea that captures all three types of associations. Various models such as repeated measures analyses, multilevel analyses, and SEM provide similar ways of capturing the associations that occur between observations. A relatively easy way to conceptualize these associations is through the covariance matrix between all observations lumping people, time, and variables. We begin with a simple description and build up the elements. Suppose we have 20 individuals measured once on a single variable and we want to estimate the mean across the 20 individuals. We can model the data as $Y_i = \mu + \varepsilon_\iota$ (i.e., a constant intercept for all 20 participants, which in this case will be the mean μ, and an error term ε). The usual assumption is that the error terms are independent and identically distributed. In other words, these 20 error terms are modeled as a 20 × 20 covariance matrix with a special structure. The diagonal contains a constant number, which is the variance of the residuals (the "identically distributed" part yields the same error variance across all observations). The off-diagonal terms are all zero because the residuals are assumed to be independent. This model imposes a theoretical covariance structure on the observations *Y* (also a 20 × 20 covariance matrix) such that any two observations are independent and there is a common variance across all observations.

It may seem like overkill to explain so much detail for the simple model of the mean of 20 observations, but this is the basic structure we need to illustrate the three types of association. The different types of association impose structure on the covariance matrix of observations, and it is helpful to take this view to gain insight into the issues surrounding interdependent data.

Now we turn to the case of temporal association by considering two observations for the same person, that is, the 20 individuals are measured twice so there are a total of 40 observations. The model for comparing the difference between the mean at each time becomes $Y_{ti} = \mu + \beta_t + \alpha_i + \varepsilon_{ti}$ with grand mean μ, a time fixed effects factor β, a subject main effect treated as a random effect factor α, and error term ε. This results in 40 error terms, which can be placed in a 40 × 40 covariance matrix.[1] The random effect terms α introduce a covariance across the

[1]In this introductory chapter, we take liberties with notation. For example, we frequently switch between two subscripts *t* and *i* to denote time (e.g., 1 or 2) and person (e.g., 1 to 20), and a single subscript *i* to denote time and person (e.g., 1 to 40). We also do not carefully distinguish population and sample parameters.

40 observations: Two observations from the same person, that is, two observations having the same α, are now associated relative to other observations even though the residuals remain independent. So, in the 40 × 40 covariance matrix of observations Person 1's Time 1 and Time 2 scores have a nonzero entry, Person 2's Time 1 and Time 2 scores have the same nonzero entry, and so on. We typically assume homogeneity of covariance, so each of the nonzero entries in the off-diagonal are constrained to be equal.

This framework can be extended to dyads. Suppose the 40 observations came from 20 heterosexual dating couples. A covariance is introduced between two members of the same dyad (interpersonal association) in just the same way as a covariance is introduced by two observations from the same person (temporal association). Similarly, the covariance between individuals from different dyads is zero, just as the covariance between observations of two different people in the case of repeated measurement is zero.

Standard methods for analysis of dyadic data automatically take proper account of the temporal, interpersonal, and multivariate associations across observations. For example, a multilevel modeling approach to dyadic data takes the information supplied by the user that two individuals are nested in the same dyad and internally constructs a covariance matrix that has the proper structure. Likewise, other approaches to dyadic data analysis, such as SEM, also establish a covariance matrix with the proper structure. It is for this reason that different approaches can be used to analyze dyadic data—the key feature is how the analytic approach structures the covariance between observations. Different analytic frameworks like multilevel models and SEMs merely become the user interface by which the user can communicate the proper structure of the covariance matrix.

In the remaining sections of this chapter, we build on this basic intuition that design features in dyadic data require particular analytic elements. Readers familiar with multilevel modeling will recognize that we treat individuals as nested within dyads. Readers familiar with SEMs will recognize that we create latent variables to model shared dyadic variance. Dyadic data analysis can be discussed in terms of either multilevel models or SEMs. We switch back and forth between both representations freely because both offer unique insights into dyadic data analysis and some problems are easier to specify in one representation than the other. The covariance representation presents a common language to discuss these different analytic strategies. We first present the simple case of a single dependent variable collected from dyad members, and then we explore other design features involving temporal and multivariate elements.

THE BASIC INTERDEPENDENT MODEL: A SINGLE DEPENDENT VARIABLE, TWO PEOPLE

A key distinction is whether dyad members are distinguishable or exchangeable (e.g., Griffin & Gonzalez, 1995). Dyad members are distinguishable when the individuals can be identified on the basis of a theoretically meaningful variable such as gender in the case of heterosexual dating couples. Dyad members are exchangeable when the individuals cannot be distinguished on the basis of a theoretically meaningful variable, such as in the case of homosexual dating couples. In this chapter we mostly focus on the distinguishable case. For details on the exchangeable case, see Griffin and Gonzalez (1995) and Kenny et al. (2006).

Interdependence between interval scaled data in the context of linear models is captured by the ICC. The basic intuition for the ICC is that it is the percentage of variance associated with between couple variance. One standard formulation takes the variance associated with a dyad-level parameter and normalizes it by the sum of that dyad-level variance plus the variance of the individual-level error term. In the context of the general linear model, the underlying structural model for an observation Y for the jth person in the ith dyad is

$$Y_{ij} = \mu + \beta_j + \alpha_i + \varepsilon_{ij}, \tag{1}$$

where the μ is the fixed effect constant, β_j is the fixed effect term for the jth subject in a dyad (such as husband and wife), α_i is a random effect for dyad that is assumed to be normally distributed with mean zero and variance, σ^2_α and the usual error term

ε with variance σ_ε^2. Note the similarity of this model and the model for the two-time repeated measures presented earlier in the chapter. There are slightly different estimation formulas for the ICC depending on whether one uses maximum likelihood or restricted maximum likelihood (the latter accounts for degrees of freedom as in an ANOVA) estimation, but the basic logic is similar. The ICC becomes the ratio

$$\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}. \quad (2)$$

Some people like to represent this framework in the context of a multilevel model with the first level representing data at the individual level and the second level representing dyads. This model is written in two parts:

$$Y_{ij} = \gamma_i + \beta_j + \varepsilon_{ij} \text{ and} \quad (3)$$

$$\gamma_i = \mu + \alpha_i, \quad (4)$$

where β is a fixed effect term that estimates, say, the difference between the two distinguishable dyad members, γ is a random effect dyad term, and the ε is the usual error term. If one substitutes Equation 4 into Equation 3, then the result is the same as Equation 1. The two approaches are the same.

A third way to conceptualize the ICC is as an SEM with two indicators, one latent factor, and a specific set of restrictions. If one sets the variance of the latent factor to one, the two indicator paths to the observed variables equal to each other, and the error variances equal to each other, then the indicator paths are equal to the square root of the ICC (see Figure 22.1). The rationale for equating the two indicator paths is because the interpretation of the latent variable is one of *shared variance* in which both individuals (regardless of whether they are exchangeable or distinguishable) contribute equally. This is a different parameterization and interpretation than the usual latent factor in which indicators can have different path estimates and so can relate differentially to the latent variable.

Thus, there are several ways to conceptualize the logic of interdependence as indexed by the ICC, and they all lead to the same result. One can model the intraclass as a linear mixed model, as a multilevel model, or as an SEM. The results will be the same as

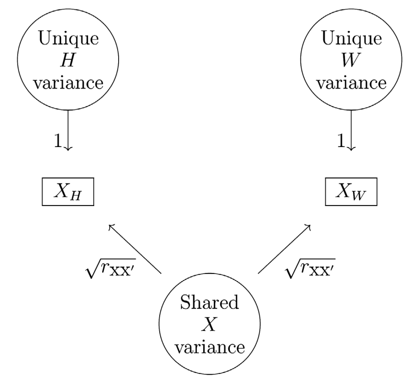

FIGURE 22.1. Illustration of the structural equations approach to estimating the intraclass correlation between husband (X_H) and wife (X_W) data.

long as the same estimation procedure is used (such as maximum likelihood) and the proper constraints on the parameters are imposed throughout. All three approaches impose the identical structure on the relevant covariance matrix. There really is no conceptual reason to favor one statistical framework over the other when it comes to dyadic data analysis. It is not necessary to limit oneself to using, say, multilevel modeling programs when one has nested data of the type seen in dyadic research. We find that other considerations, such as the ability to relax some constraints or the ability to handle missing data in sensible ways, turn out to be important in choosing one statistics package or modeling framework over another. For example, it may be easier to test some constraints, such as the equality of error variance across individuals, or to generalize the model to allow for differential error variance in one approach rather than another. The interdependence issue present in dyadic data analysis does not by itself force an analyst into one specific type of statistical representation.

Throughout this chapter, we focus on a normally distributed interval scale data. Given that we operate in the context of a linear mixed model (i.e., a general linear model with random effects), it is relatively straightforward to test these models within the generalized linear mixed model (GLMM). This generalization includes extensions to regression models on various distributions such as the binomial, the negative binomial, and the Poisson. These extensions may turn out to be relatively straightforward for simple models, but as we introduce more complicated multivariate dyadic models, the GLMM approach needs to be studied more carefully given that some distributions impose some challenging restrictions on some parameters, for example, the definition of the ICC in the context of binomial data is tricky because the usual GLMM logit link function imposes a constraint on the error variance (e.g., Snijders & Bosker, 1999). The generalization of dyadic models to nonnormal distributions remains an open area of research.

A MULTIVARIATE INTERDEPENDENT MODEL: TWO DEPENDENT VARIABLES, TWO PEOPLE

Latent Variable Model

Our discussion of the ICC so far has focused on one variable (Figure 22.1). We extend this framework to the case of two variables observed on each member of the dyad (Figure 22.2). For instance, we collect data on depression and marital satisfaction for each member of the couple. In addition to two ICCs, one for each variable *X* (depression) and *Y* (satisfaction), the model adds two new terms that span the two variables. One term is the dyad-level covariance, which is interpreted as the covariance between two latent variables, or the dyadic relation between dyadic depression and dyadic satisfaction. In the context of dyadic data analysis, the latent variable estimated under the constraints shown in Figure 22.2 represents the shared variance between husband and wife within each respective variable. So the covariance is the covariance between the shared variance on depression and the shared variance on marital satisfaction. The dyad correlation is not equivalent to the correlation between the dyad means (Griffin & Gonzalez, 1995). The individual-level covariance is the covariance between the individual factor variance on each variable. That is, the husband and wife each have error variance remaining after accounting for the shared dyadic variance on each variable. The individual-level covariance represents the covariance across those individual-level error terms (e.g., the covariance between the husband's individual error term on depression and his individual error term on marital satisfaction). When couple members are distinguishable, it is possible to test the assumption of equal individual-level covariances by allowing the two terms to be freely estimated and comparing the free model to the constrained model using the likelihood ratio test (i.e., Gonzalez & Griffin, 2001).

The model depicted in Figure 22.2 can be equivalently estimated in the context of a multilevel model. Unfortunately, the description of the multivariate

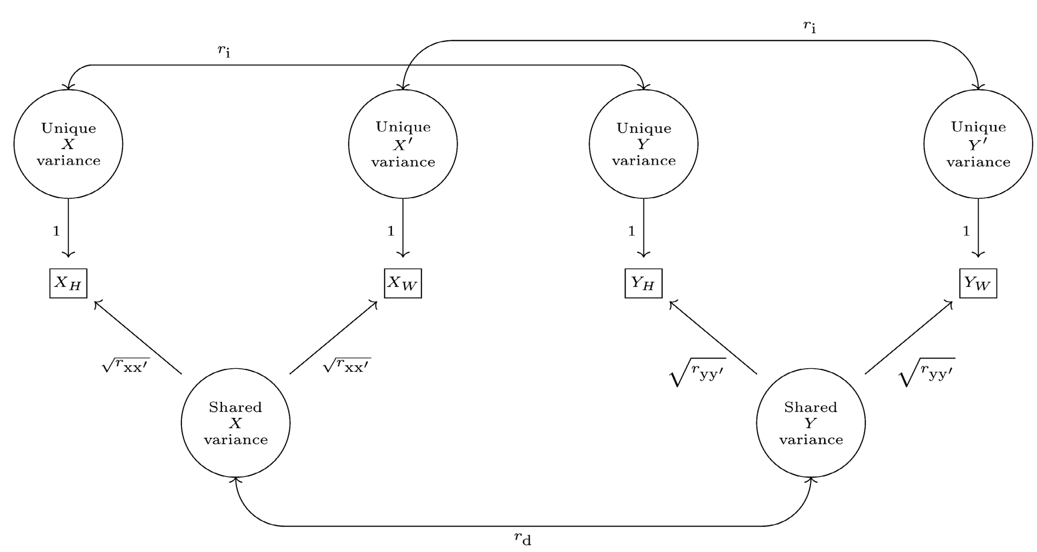

FIGURE 22.2. Illustration of the structural equations approach to estimating the intraclass correlation between husband and wife on two variables *X* and *Y*, the dyad level correlation r_d and the individual level correlation r_i.

model is not straightforward because multilevel models are typically expressed for a single variable with a single residual term ε at level 1. But we need separate error terms for each of the two variables. As we show later in this section, to fit the multilevel model to the multivariate example in Figure 22.2, it is necessary to express the multilevel model in a way that allows for different residual variances for each variable. We present one method to model multiple variables in the multilevel model using a switching regression technique.

The variables are represented in a single long column of data, which we denote **Y**. That is, we have a single column of data that includes the husband's depression, the wife's depression, the husband's marital satisfaction, and the wife's marital satisfaction. So, if there are 20 dyads, there are a total of 80 observations because each dyad member contributes two scores. It may seem strange to place data from different people and different variables into the same column, but by using proper codes, we can account for data associated with different variables and different people. The approach constructs the appropriate 80 × 80 covariance matrix for the 80 observations. One way to account for the dependence of multiple variables is to use the repeated measures option common in most multilevel model programs (e.g., depression and marital satisfaction would be treated as a repeated measures with an unconstrained covariance matrix, including heterogeneous variances, because we would not necessarily require that the variance for depression equal the variance for marital satisfaction). Then, the dyadic part of the analyses is implemented in a two-level model. This approach is described in Kenny et al. (2006).

An equivalent way to represent this multivariate framework in a multilevel model involves a *switching regression* approach, and we briefly review it here because it illuminates key points about the multivariate dyadic latent model (see Gonzalez & Griffin, 2001). First, place all observations from both partners and from both variables into a single long column **Y**. Second, create two columns of dummy codes. One dummy code *D*1 assigns a one to all depression scores and a zero for all marital satisfaction scores. The other dummy code *D*2 assigns a zero to all depression scores and a one to all marital satisfaction scores. The dummy codes perfectly partition the long column of observations into depression data and into marital satisfaction data. The first-level equation is written as follows and does not include an error variance at Level 1 (this can be easily implemented in the MLWin multilevel modeling program) nor an intercept:

$$\mathbf{Y} = \beta_d D1 + \beta_S D2. \tag{5}$$

These two βs from Level 1 are modeled as random effects. The error terms for each variable can be modeled separately for each variable because they will be attached to different βs at the second level. We omitted the intercept and the error term at this first level to estimate them separately for each variable at the higher level.

The second level of regression equations are as follows:

$$\beta_d = \text{intercept}_\text{d} + v_\text{dg} + u_\text{di} \text{ and} \tag{6}$$

$$\beta_s = \text{intercept}_s + v_{sg} + u_{si}, \tag{7}$$

where *v* and *u* are random effects that code group and individual terms (respectively), each equation has its own fixed effect intercept term, the subscripts *d*, *s*, *g*, and *i* refer to depression, satisfaction, group, and individual, respectively. The random effect *v* assesses group-level variance, and the random effect *u* assesses individual-level variance. In short, the switching regression (Level 1) isolates the two variables depression and marital satisfaction, and the next two levels capture the dyadic structure. This is implemented as a three-level model as far as the statistical program is concerned, but some researchers would call this a two-level model given that there is no error variance in the first level with the switching regression (Equation 5).

Now comes the important part of this particular formulation, which provides some new intuition. We formulate a covariance structure on each of the random effects *v* (group level) and *u* (individual level). Let the two group-level *v*s be bivariate normally distributed with covariance matrix

$$\Omega_v = \begin{bmatrix} \sigma^2_{vd} & \\ \sigma_{vds} & \sigma^2_{vs} \end{bmatrix}, \tag{8}$$

where *d* and *s* denote depression and satisfaction, respectively. This means that the random effect *v* associated with depression has variance σ^2_{vd}, random effect *v* associated with satisfaction has variance σ^2_{vs}, and the two have covariance σ_{vds}. Similarly, an analogous covariance is formulated on the two individual-level *u*s

$$\Omega_u = \begin{bmatrix} \sigma^2_{ud} & \\ \sigma_{uds} & \sigma^2_{us} \end{bmatrix}. \quad (9)$$

This covariance matrix gives the variances and covariance between depression and satisfaction at the individual level. In this formulation, we require equality of all individual-level correlations (i.e., referring to Figure 22.2, this particular multilevel model implementation forces the two individual-level correlations for husband and wife to be identical).

These two covariance matrixes contain information about group-level and individual-level variance for each variable and information about group-level and individual-level covariance between the two variables. They provide all the information necessary to compute the terms in the latent group model as well as each of the two ICCs. Using the terms in those two covariance matrixes, we have

$$\text{intraclass correlation for depression: } \frac{\sigma^2_{vd}}{\sigma^2_{vd} + \sigma^2_{ud}},$$

$$\text{intraclass correlation for satisfaction: } \frac{\sigma^2_{vs}}{\sigma^2_{vs} + \sigma^2_{us}},$$

$$\text{individual level correlation between depression and satisfaction: } \frac{\sigma_{uds}}{\sqrt{\sigma^2_{ud}\sigma^2_{us}}}, \text{ and}$$

$$\text{dyad level correlation between depression and satisfaction: } \frac{\sigma_{vds}}{\sqrt{\sigma^2_{vd}\sigma^2_{vs}}}.$$

The two intraclass definitions are identical to what we presented in a previous section. The form of the individual- and dyad-level correlations is the usual correlation (a covariance divided by the square root of a product of variances). The individual-level correlation uses terms from the individual-level covariance matrix *u* and the group-level correlation uses terms from the group-level covariance matrix *v*. Thus, these two covariance matrixes yield the ICCs, the variances of the individual- and group-level latent variables, and the individual- and group-level correlations, and the matrixes are identical to the SEM represented in Figure 22.2. This framework sets up the same structure on the 80 × 80 covariance matrix as the SEM in Figure 22.2.

The basic dyadic structure for the two variable latent model can be extended to more variables and to more complicated models. For example, one can take the two-variable model described in Figure 22.2 and use those variables (depression and marital satisfaction) to predict a third variable, say, parenting quality. The prediction can be modeled at both the individual level and the dyad level so that the prediction of parenting quality can occur at two levels. Does the shared variance of depression (dyad level) predict parental quality? Does the shared variance of marital satisfaction predict parental quality? Does the individual-level variance of the husband's depression predict his individual-level variance of parenting?

The take-home message of this type of modeling is that the variance and covariance is partitioned into individual-level and dyad-level terms. Once the partition is performed, many research questions can be tested at each level. It is in this way that the analysis of dyadic data goes beyond "correcting" data for violations of independence and instead directly modeling the interdependence. We can see how partitioning the analysis into separate individual-level and dyad-level covariance matrixes provides useful terms, such as ICCs, to illuminate the study of interpersonal processes.

Actor–Partner Model

A different representation of the latent variable model in Figure 22.2 is the *actor–partner model* (APM) depicted in Figure 22.3. The difference between the two models is how the interdependence is modeled—the latent variable model uses latent variables, whereas the APM uses observed variables directly. Is the wife's marital satisfaction predicted by both the husband's and the wife's depression? In this research question, there are no latent variables representing dyadic- and individual-level variance. Instead, we have two actor paths (*a* and *d*) that represent an individual's influence of depression on the same individual's score of marital satisfaction. There are two cross-person cross-variable paths (*b* and *c*)

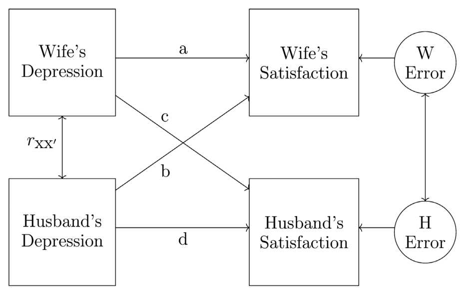

FIGURE 22.3. Actor–partner model for depression and marital satisfaction.

that represent the effect of wife's depression on husband's marital satisfaction and the effect of husband's depression on wife's satisfaction. The two exogenous predictors (depression from each couple member) are correlated and the residuals associated with satisfaction as the dependent variable are correlated. The model fits from the latent variable model in Figure 22.2 and the APM in Figure 22.3 are equivalent under some special restrictions. When two equality constraints are placed on the path coefficients ($a = d$ and $b = c$), the two residual variances are constrained to be equal, and the variances of the two predictors (depression) are constrained to be equal, then the models yield identical measures of fit. Of course, it is possible to generalize the models in different ways, and this is what makes each of them differentially useful as models. In general, a good use of the latent variable model formulation is when one wants to partition individual- and dyad-level variance; a good use of the APM is when one wants to examine interrelations of individuals (in particular, how one couple member's data predicts the other member's data). Research questions involving similarity of dyad members are naturally tested with the latent variable model, whereas research questions of interpersonal influence are naturally tested with the APM (Gonzalez & Griffin, 2001). In some literatures the APM is known as a *seemingly unrelated regression* (SUR).

THE LONGITUDINAL INTERDEPENDENT MODEL: A SINGLE VARIABLE, TWO TRAJECTORIES, TWO PEOPLE

There has been much progress on longitudinal data analysis, such as latent growth curve analysis (e.g., Singer, 1998). We can combine advances from latent growth modeling with those in dyadic models. Within our framework, these hybrid designs produce data with both temporal and interpersonal association. For simplicity in exposition, we assume a relatively simple example in which the investigator has three time-points, and each member of the couple provides depression data at each of those three time-points. There are at least two ways one can approach the modeling of this situation. One is to focus on the latent growth curve to model a curve for each person. This places priority on the temporal association. To deal with the interpersonal association, or interdependence, one could allow covariances across key terms in the latent growth curve. Suppose we fit both a random intercept and a random slope to the three waves of depression data with separate trajectories for the wife and for the husband. To model the interdependence attributable to couples, we estimate additional covariances between the husband and wife's random effect intercept, a covariance between the husband and wife's random effect slope, and cross-covariances spanning one person's intercept with the other's slope. The model assesses, for instance, the relation between latent parameters of the growth model, the intercept and slope, across dyad members.

A second way to model these data is to take the latent variable dyadic model as primary and apply it three times, once for each time-point. So there are three V-style diagrams as in Figure 22.1, one for each time the depression variable is assessed. This portion of the model accounts for the interpersonal association. To account for temporal association, one can estimate covariance terms across individual- and dyadic-level variances, or instead of covariances, one could estimate an autoregressive model having regression paths across time predicting the individual- and dyad-level variances. This model can be used to assess the stability of the dyadic- and individual-level variances.

These two model frameworks differ in their priority. The first model places priority on the trajectories over time and addresses interdependence over the parameters of the trajectory. The second model places priority on the interdependence between dyad members and addresses temporal association over the interdependence parameters. Both models can yield identical model fits depending on how they are parameterized and the nature of constraints that are imposed. That is, they both can imply the same covariance structure on the observations by properly accounting for both temporal and interpersonal association in the observations. They can also be relaxed and generalized in different ways and so can be used to test different models with different psychological implications. Both of these frameworks can be specified in the context of multilevel models. For an example of a longitudinal design in the multilevel framework for couple research using a single dependent variable, see Barnett, Raudenbush, Brennan, Pleck, and Marshall (1995). We prefer to conceptualize latent growth models in the context of SEM because it is sometimes easier to generalize the temporal part of the model, such as being able to specify unconstrained error variances at each time. But we recognize that both the multilevel and structural equation modeling approaches to latent growth modeling have their merits and can be identical depending on how the constraints are imposed within each framework.

A LONGITUDINAL, MULTIVARIATE INTERDEPENDENT MODEL: TWO DEPENDENT VARIABLES, TWO TRAJECTORIES, TWO PEOPLE

This design contains all three types of association: temporal, multivariate, and interpersonal. It arises when the research design measures couples on multiple variables such as depression and marital satisfaction over multiple time-points. For example, husband and wife dyad pairs are assessed on both depression and marital satisfaction at each of three time-points. We describe one way to model these three associations simultaneously. Conceptually, we can estimate the latent group model (i.e., the model depicted in Figure 22.2) separately at each time. But rather than estimating three separate submodels (one for Time 1, a second for Time 2, and a third for Time 3), it is appropriate to estimate the entire model simultaneously (i.e., the latent group model at all three times together). This simultaneous model permits, among other things, the estimation of stability of the latent variable terms. This can be accomplished either by estimating covariances between all time i latent variables and all time $i + 1$ latent variables or by estimating regression paths between two adjacent time-points for different terms in the multivariate interdependent model. There is a sense that this representation gives priority to interdependence, then models the cross-variable association and the temporal association.

Other frameworks are possible too, such as placing priority on the temporal part and estimating trajectories for each person on each variable. Once the latent trajectories are estimated, then the multivariate and interpersonal associations are added across the relevant trajectory parameters. One can also perform additional analyses (such as moderation and mediation) to gain a deeper understanding of the contributors to the different elements of this general model, such as examining the predictors of group-level stability or individual-level stability, or examining interesting combinations such as cross-variable individual-level stability, for example, examining the correlation between the error for husbands' depression at Time 1 with the error of husbands' marital satisfaction at Time 2. This model can also be implemented in a multilevel framework.

STRUCTURAL EQUATION MODELING, MULTILEVEL MODELS, AND GENERALIZATIONS

Our general formulation of temporal, multivariate, and interpersonal association has been useful for organizing many of the dyadic models that have been discussed in the literature. We see that different implementations, such as in an SEM approach or in a multilevel modeling approach, offer advantages and disadvantages. Sometimes it is easier to work with an SEM program such as when dyads are distinguishable, there are multiple variables, or one wants to test mediation and moderation. Sometimes a multilevel model approach is easier to implement, such as when dyads are exchangeable, there are missing data, or one is primarily concerned with partitioning dyad and individual random effect variance. Some generalizations or tests of constraints are easier within one framework than another, but many features can be implemented with either approach. For example, exchangeable and distinguishable dyads can be modeled with either multilevel models or with SEMs (e.g., Woody & Sadler, 2005).

The analysis of the latent dyad model can be conducted even when some groups have data from only one member (e.g., Snijders & Bosker, 1999). Multilevel techniques permit *units* of unequal sizes. So, if the unit is the individual with, say, four time-points, then one way to handle missing data would be to treat the missingness as something that yields unequal-size units. The general multilevel framework can thus handle missing observations in a longitudinal design as well as groups of unequal sizes (as would be encountered, for example, in research in which a family is the unit of analysis and families vary in size); for a complete discussion, see Raudenbush and Bryk (2002).

We view the close connection to both structural equation and multilevel modeling frameworks as a plus rather than a negative, as providing insight rather than a source of confusion. As new methods become available in either framework, the dyadic researcher can easily switch back and forth to take advantage of new developments. For example, recent advances in growth mixture modeling techniques for simultaneously estimating trajectories and identifying different classes of individuals exhibiting different trajectory patterns could be extended to dyadic research.

SOFTWARE IMPLEMENTATIONS

The dyadic researcher has a broad array of software choices. Major statistical programs (such as SPSS, SAS, R, and Stata) can run the procedures described in this chapter. The user must be mindful of the particular defaults in any statistical program they use. Researchers preferring an SEM framework can use any SEM program, including EQS, LISREL, AMOS, and Mplus. Those who prefer a multilevel modeling approach can use programs such as HLM, MLWin, and Mplus. Again, careful attention to defaults in any statistical program one uses is important.

Some statistical programs such as Mplus and SAS NLMIXED make estimating and testing general models relatively easy. We welcome such flexible software. At the same time, we caution researchers that as new procedures are created by mixing and matching diverse elements (such as a dyadic longitudinal growth mixture model on binary data), it becomes even more important to understand the statistical underpinnings of one's model. It is common to see strange behavior in such novel hybrid models, such as negative error variances. The researcher should seek methodological advice when exploring new modeling territory. It is not advisable to use quick fixes, such as setting a negative error variance to zero, or setting a particular equality constraint, merely because that makes the program converge or run without error. Rather, when a new hybrid model produces strange output, it is best to consult a methodologist.

We recommend that the dyadic researcher maintain an open mind when choosing among different statistical packages. There is probably no single statistics package, or framework, that will serve all research needs. The dyadic researcher will also need to be flexible in setting up the data in different ways for different analyses. Sometimes dyadic analyses can be done in wide format, such that each row represents a couple, and data from both members of the couple appear on the same row of the data matrix (as when using an SEM framework). Other times analyses are easier in long format, such that each row represents a person, so data from a couple spans two rows in the data matrix, as when using a multilevel framework. Each of those two data formats requires careful use of couple codes; data for distinguishable individuals is coded by individual codes in the long format but automatically by columns in the wide format. Sometimes data need to be entered twice, in the sense that one column contains data from Harry and Sally (separate rows), whereas a second column contains the same data for the same variable but in reverse order (i.e., Sally and Harry). This type of data entry can be used for the pairwise approach to dyadic data analysis (Griffin & Gonzalez, 1995) and in the APM (Kenny et al., 2006), and it directly provides maximum likelihood estimates in the case of no missing data.

CONCLUSION

Dyadic designs provide researchers an exciting route to study interpersonal processes. The inclusion of both individuals from the dyad in the data set provides the opportunity to study interdependence. Interdependence is not a problem with the data but is a limitation of the standard statistical techniques psychologists tend to use. By framing interdependence in the language of associations, we showed that it is possible to discuss temporal, interpersonal, and multivariate associations together. The richness of dyadic data analysis is such that sometimes it is longitudinal, sometimes multivariate, and sometimes both. So it is necessary to take into account three sources of association that may be present in a dyadic design. We hope this chapter has convinced the reader that it is no longer necessary to avoid analyzing interdependent data that violate the independence assumption and has inspired the reader to study dyadic processes.

References

Barnett, R. C., Raudenbush, S. W., Brennan, R. T., Pleck, J. H., & Marshall, N. L. (1995). Change in job and marital experiences and change in psychological distress: A longitudinal study of dual-earner couples. *Journal of Personality and Social Psychology, 69*, 839–850. doi:10.1037/0022-3514.69.5.839

Card, N., Selig, J., & Little, T. (2008). *Modeling dyadic and interdependent data in the developmental and behavioral sciences.* New York, NY: Routledge.

Gonzalez, R., & Griffin, D. (2001). A statistical framework for modeling homogeneity and interdependence in groups. In M. Clark & G. Fletcher (Eds.), *Handbook of Social Psychology: Vol. 2. Interpersonal processes* (pp. 505–534). Malden, MA: Blackwell.

Griffin, D., & Gonzalez, R. (1995). The correlational analysis of dyad-level data: Models for the exchangeable case. *Psychological Bulletin*, *118*, 430–439. doi:10.1037/0033-2909.118.3.430

Kenny, D. A. (1990). Design issues in dyadic research. In C. Hendrick & M. S. Clark (Eds.), *Review of Personality and Social Psychology: Vol. 11. Research methods in personality and social psychology* (pp. 164–184). Newbury Park, CA: Sage.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in the analysis of variance. *Psychological Bulletin*, *99*, 422–431. doi:10.1037/0033-2909.99.3.422

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis.* New York, NY: Guilford Press.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models and individual growth models. *Journal of Educational and Behavioral Statistics, 23*, 323–355.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage.

Woody, E., & Sadler, P. (2005). Structural equation models for interchangeable dyads: Being the same makes a difference. *Psychological Methods, 10*, 139–158. doi:10.1037/1082-989X.10.2.139