

DIARY METHODS: Capturing Life as it is Lived

Niall Bolger, Angelina Davis, and Eshkol Rafaeli

Psychology Department, New York University, New York, New York 10003;
e-mail: niall.bolger@nyu.edu, aed232@nyu.edu, erm5@nyu.edu

Key Words experience sampling method, longitudinal designs, electronic data collection, self-report measures, multilevel models

■ **Abstract** In diary studies, people provide frequent reports on the events and experiences of their daily lives. These reports capture the particulars of experience in a way that is not possible using traditional designs. We review the types of research questions that diary methods are best equipped to answer, the main designs that can be used, current technology for obtaining diary reports, and appropriate data analysis strategies. Major recent developments include the use of electronic forms of data collection and multilevel models in data analysis. We identify several areas of research opportunities: 1. in technology, combining electronic diary reports with collateral measures such as ambulatory heart rate; 2. in measurement, switching from measures based on between-person differences to those based on within-person changes; and 3. in research questions, using diaries to (a) explain why people differ in variability rather than mean level, (b) study change processes during major events and transitions, and (c) study interpersonal processes using dyadic and group diary methods.

CONTENTS

INTRODUCTION	580
DIARY STUDY DESIGN	581
A Typology of Research Questions	581
Types of Diary Designs	588
Limitations in the Use of Diaries	591
Design Section Summary	593
DIARY STUDY TECHNOLOGY	593
Paper and Pencil Diaries	593
Augmented Paper Diaries	595
Handheld and Electronic Data Collection	596
Emerging Opportunities	599
Technology Section Summary	599
DIARY DATA ANALYSIS	600
Aggregating Over Time: What is the Typical Person Like, and How Much Do People Differ from Each Other?	601
Modeling the Time Course: How Does a Typical Person Change Over Time, and How Do People Differ in Change Over Time?	604

Modeling Within-Person Processes: What is the Within-Person Process for the Typical Person, and How Do People Differ in These Processes?	606
Other Data Analysis Issues	608
Analysis Section Summary	610
CONCLUSIONS	610

INTRODUCTION

Gordon Allport, writing in 1942, pointed out that an acquaintance with the “particulars of life” is the beginning of all psychological knowledge—scientific or otherwise. “Psychology needs to concern itself with life as it is lived, with significant total-processes of the sort revealed in consecutive and complete life documents” (Allport 1942, p. 56). This chapter reviews the state of the art in research answering Allport’s call, broadly defined today as diary methods. Methods for documenting the particulars of life have improved considerably over the years. Diaries, self-report instruments used repeatedly to examine ongoing experiences, offer the opportunity to investigate social, psychological, and physiological processes, within everyday situations. Simultaneously, they recognize the importance of the contexts in which these processes unfold. Thus, diaries are designed to capture the “little experiences of everyday life that fill most of our working time and occupy the vast majority of our conscious attention” (Wheeler & Reis 1991, p. 340).

A fundamental benefit of diary methods is that they permit the examination of reported events and experiences in their natural, spontaneous context, providing information complementary to that obtainable by more traditional designs (Reis 1994). Another is the dramatic reduction in the likelihood of retrospection, achieved by minimizing the amount of time elapsed between an experience and the account of this experience. Diaries provide the field of psychology with a powerful set of methods for studying various human phenomena, including personality processes (e.g., Bolger & Zuckerman 1995, Fabes & Eisenberg 1997, Rhodewalt et al. 1998), marital and family interaction (e.g., Almeida et al. 1999, Downey et al. 1999, Repetti & Wood 1997), physical symptoms (e.g., Suls et al. 1994), and mental health (e.g., Alloy et al. 1997). Wheeler & Reis (1991), and more recently Reis & Gable (2000), have provided comprehensive reviews of multiple domains in which diary methods have been used. Edited volumes, summarizing diary research in particular domains [e.g., psychopathology (deVries 1992)] have begun appearing. Recently, Stone & Shiffman (2002) proposed a set of guidelines for the reporting of diary studies. Rather than overlaps with these authors, we provide an overview of the current and potential uses of diaries, and we call attention to some of the limitations of this method. We do so in three major sections.

The design section reviews the types of questions appropriately addressed using diary methods, presents an overview of the main designs, and draws attention to the limitations of such designs. The technology section reviews the evolution of

measurement and instrumentation in diary research and discusses the benefits and costs of each successive innovation. Finally, the analysis section presents the basic analytic concepts that should be considered by a researcher before conducting diary research. Throughout the chapter, we provide examples of research questions from the field of adult relationships. These examples are designed to illustrate the breadth of information that can become available when diary methods are used to study a particular domain.

DIARY STUDY DESIGN

A Typology of Research Questions

The effectiveness of diary study designs depends on careful consideration of the question(s) one seeks to answer. A poorly designed diary study can involve considerable effort but may yield little useful information. Three broad types of research goals can be achieved using diary designs: (a) obtaining reliable person-level information; (b) obtaining estimates of within-person change over time, as well as individual differences in such change; and (c) conducting a causal analysis of within-person changes and individual differences in these changes.

AGGREGATING OVER TIME: WHAT IS THE TYPICAL PERSON LIKE, AND HOW MUCH DO PEOPLE DIFFER FROM EACH OTHER? It is common in psychological research to ask participants to retrospect over weeks and months and provide summary accounts of their psychological states and experiences. Depression symptom scales ask for retrospection over 14 days (e.g., the Center for Epidemiologic Studies Depression Scale; Radloff 1977). Life event inventories ask for retrospection over six months to a year (Dohrenwend et al. 1978). The perils of retrospection have been reviewed elsewhere (e.g., Nisbett & Wilson 1977, Ross 1989, Tourangeau et al. 2000). Diary data can be used to generate summary accounts without the biases introduced by retrospection over relatively long periods. In addition to yielding estimates of within-person central tendency, the diary approach can show how much people vary over time in variables of interest. The value of diary methods for such descriptive purposes has been emphasized by several prominent researchers (e.g., Affleck et al. 1999, Tennen et al. 1991).

We begin our discussion with a hypothetical study that uses aggregation. A researcher is interested in investigating women's perceptions of intimacy in their romantic relationships in everyday life. The researcher wishes to assess (a) the degree of intimacy the typical woman in a committed relationship feels on average, (b) the extent to which the typical woman's feelings of intimacy vary over time, and (c) whether women differ from one another in their average feelings of intimacy and in the variability of their feelings of intimacy over time (see Table 1). The first step is to determine an appropriate diary design.

Even when diaries are used solely to obtain aggregate measures, researchers must determine the frequency and duration of assessments that are appropriate for the phenomenon under study. If the researcher believes intimacy could vary from

TABLE 1 Examples of research questions, design, technology, and statistical analysis for time-based diary studies

I. Aggregating over time (Intervention Study; see Figure 1)			
Abstract research questions	What is the typical person's (a) average level of Y? (b) variability in Y?	What are the between-person differences in (a) average level of Y? (b) variability in Y?	What are the sources of between-person differences in (a) average level of Y? (b) variability in Y?
Concrete research questions: intimacy examples	What is the typical woman's (a) average level of intimacy? (b) variability in intimacy?	How much do women differ in their (a) average level of intimacy? (b) variability in intimacy?	How much do women in the intervention group differ from those in the control group in their (a) average level of intimacy? (b) variability in intimacy?
Design	Time-based diary completed weekly for 16 weeks	Time-based diary completed weekly for 16 weeks	Time-based diary completed weekly for 16 weeks
Technology	PDA or paper and pencil	PDA or paper and pencil	PDA or paper and pencil
Statistical model (multilevel)			
Name	One-Way Random Effects ANOVA ^a	One-Way Random Effects ANOVA	Means-As-Outcomes Model
Specification	Level 1: intercept Level 2: intercept	Level 1: intercept Level 2: intercept	Level 1: intercept Level 2: intercept, group
Relevant parameters	Average L-1 (a) intercepts (b) error variance ^b	Variance of L-1 (a) intercepts (b) error variances	Group differences in L-1 (a) intercepts (b) error variances
Abstract research questions	What is the typical person's (a1) starting level of Y? (a2) rate of change in Y? ^c (b) residual variability in Y?	What are the between-person differences in (a1) starting level of Y? (a2) rate of change in Y? (b) residual variability in Y?	What are the sources of between-person differences in (a1) starting level of Y? (a2) rate of change in Y? (b) residual variability in Y?

II. Modeling time (Intervention Study; see Figure 1)

Concrete research questions: intimacy examples	What is the typical woman's (a1) starting level of intimacy? (a2) rate of change in intimacy? (b) residual variability in intimacy?	How much do women differ in their (a1) starting level of intimacy? (a2) rate of change in intimacy? (b) residual variability in intimacy?	Do women in the intervention group differ from those in the control group in their (a1) starting level of intimacy? (a2) rate of change in intimacy? (b) residual variability in intimacy?
Design	Time-based diary completed weekly for 16 weeks	Time-based diary completed weekly for 16 weeks	Time-based diary completed weekly for 16 weeks
Technology	PDA or paper and pencil	PDA or paper and pencil	PDA or paper and pencil
Statistical model (multilevel) Name	Linear Growth Model	Linear Growth Model	Linear Growth Model with Level-2 Predictor
Specification	Level-1: intercept, time Level-2: intercept	Level-1: intercept, time Level-2: intercept	Level-1: intercept, time Level-2: intercept, group
Relevant parameters	Average L-1 (a1) intercept (a2) time slope (b) error variance	Variance of L-1 (a1) intercept (a2) time slopes (b) error variances	Group differences in L-1 (a1) intercept (a2) time slopes (b) error variances

III. Modeling within-person process^d (Daily Conflicts Study; see Figure 2)

Abstract research questions	What is the typical person's (a1) level of Y when X does not occur (is 0)? ^c (a2) change in Y when X occurs (when X differs by one unit)? (b) residual variability in Y?	What are the between-person differences in (a1) level of Y when X does not occur? (a2) change in Y when X occurs? (b) residual variability in Y?	What are the sources of between-person differences in (a1) level of Y when X does not occur? (a2) change in Y when X occurs? (b) residual variability in Y?
--------------------------------	--	---	--

(Continued)

TABLE 1 (Continued)

Concrete research questions: intimacy examples	What is the typical woman's (a1) level of intimacy on days when no conflicts occur? (a2) difference in intimacy between conflict and no-conflict days? (b) residual variability in intimacy?	How much do women differ in their (a1) level of intimacy on days when no conflicts occur? (a2) difference in intimacy between conflict and no-conflict days? (b) residual variability in intimacy?	Do women low in Relationship Quality (RQ) differ from those high in RQ in their (a1) level of intimacy on days when no conflicts occur? (a2) difference in intimacy between conflict and no-conflict days? (b) residual variability in intimacy?
Design	Time-based diary completed daily for 28 days	Time-based diary completed daily for 28 days	Time-based diary completed daily for 28 days
Technology	PDA or paper and pencil	PDA or paper and pencil	PDA or paper and pencil
Statistical model (multilevel)			
Name	Random-Coefficients Regression Model	Random-Coefficients Regression Model	Intercepts-and-Slopes-as-Outcomes Model
Specification	Level 1: intercept, conflict day Level 2: intercept	Level 1: intercept, conflict day Level 2: intercept	Level 1: intercept, conflict day Level 2: intercept, RQ
Relevant parameters	Average L-1 (a1) intercept (a2) conflict day slope (b) error variance	Variance of L-1 (a1) intercepts (a2) conflict day slopes (b) error variances	RQ differences in L-1 (a1) intercepts (a2) conflict day slopes (b) error variances

^aNames of statistical models are taken from Raudenbush & Bryk (2002); ^bResearch questions about level-1 variances require models that allow for heterogeneous error variances; ^cFor simplicity, we assume linear rates of change with time; ^dAlthough not shown here, process models should include controls for any time trends; ^eFor simplicity, we describe a cross-sectional analysis. For details on a longitudinal analysis, see text.

day to day, but can be recalled quite well over the span of a day, he or she may choose a fixed assessment schedule in which participants provide reports at fixed times (e.g., daily for four weeks). Obtaining sufficient diary entries will enable the researcher to address questions not only about between-person differences in feelings of intimacy but also about between-person differences in the variability of feelings of intimacy. Between-person differences in variability is an important but neglected topic in psychological research [but see the work of Kernis and colleagues for a notable exception (e.g., Greenier et al. 1999)].

The traditional alternative to using aggregated diary data has been the use of single reports in which participants attempt to recall their experience. Such retrospection is often plagued by biases. Participants' limited ability to recall often results in retrospective "aggregate" responses that reflect faulty reconstruction of the phenomena of interest. Retrospection is susceptible to state-congruent recall: The current state cues similar or similarly valenced instances, leading to a biased report (e.g., Bower 1981). A related finding is that respondents, at least when reflecting on pain, tend to rely on a "peak-end" rule, giving more weight to the peak levels and to the most recent levels of experience rather than equally weighting each instance (Redelmeier & Kahneman 1996) or day (Stone et al. 2000). Shiffman et al. (1997) demonstrated that subjective aggregates have a poorer fit to actual (diary-reported) experiences, than do empirical aggregates based on these responses. Thus, one advantage of diary-based data, even for such simple questions, lies in the reduction in systematic and random sources of measurement error, and with it the increase in validity and reliability.

Still, a researcher uninterested in modeling time or other within-person factors may question the need for a diary study. Indeed, it is unlikely that a diary design will be used solely in a study of this kind. However, seemingly simple questions such as these can be posed and answered within more elaborate diary designs, and they should be examined before moving on to more complex ones (detailed below). For example, if within-person variability in a measure is small, there is little point in pursuing within-subject research questions. Also, in such cases it is more likely that retrospective accounts will be accurate.

MODELING THE TIME COURSE: HOW DOES A TYPICAL PERSON CHANGE OVER TIME, AND HOW DO PEOPLE DIFFER IN CHANGE OVER TIME? In addition to their utility in answering questions that do not directly involve time as a factor, diary designs are excellent for studying temporal dynamics. By having participants report their experiences over hours, days, weeks, and sometimes months, researchers can ask questions such as: Does the variable of interest fluctuate from morning to night, behave differently on weekends and weekdays, or have a certain growth visible over weeks or months? Do individuals differ in these changes over time? If so, what explains these varying time courses? Traditional longitudinal designs can also address these questions, but because they typically involve only a small number of repeated measurements taken at long intervals, they cannot capture changes with the same fidelity. Of course, diary designs place a greater burden on

participants, and as we will argue later, such designs are best implemented at times when change is likely to occur.

Consider examples of how incorporating time can augment our understanding of intimacy, its average, and its variability. Intimacy is defined as an individual's feelings of being understood, validated, and cared for by another individual (Reis & Shaver 1988). A researcher may be interested in whether intimacy levels are stable or variable during a person's typical day. The diary design discussed earlier obtained intimacy ratings only once per day. For the present purpose, multiple intimacy ratings will be necessary within each day, using a fixed assessment schedule (e.g., every three waking hours, or every eight hours to represent morning, afternoon, and evening) for a period lasting two or three weeks.

Consider a different research example. An investigator is interested in temporal patterns in a young mother's intimacy with her spouse following the birth of their first child. Existing longitudinal research (e.g., Hackel & Ruble 1992) shows that levels of marital satisfaction often decline after the birth of the first child. Using weekly diaries over several months, it would be possible to determine when and to what extent the decline occurs for the average mother. Moreover, individual differences in this pattern (i.e., in both the level of intimacy reported shortly after giving birth and the effect of elapsed time on intimacy reported in the subsequent months) can also be assessed. These individual differences may be accounted for by person- or couple-level characteristics such as prenatal relationship satisfaction or length of the relationship with the spouse.

At times, researchers may be interested in cyclical, rather than linear, changes over time. Diary data have been used to examine circadian or diurnal rhythms (e.g., Clark et al. 1989, Dermer & Berscheid 1972, Rafaeli & Reville 2002), weekly cycles (e.g., Larsen & Kasimatis 1990), as well as longer time-units [e.g., menstrual cycles (cf. van Goozen et al. 1997), and seasons (Reardon et al. 2001)]. The cyclical nature of such data can be modeled using sine/cosine curves; procedures such as spectral analysis (Larsen 1990) are used to obtain the various parameters of these curves (length, amplitude, phase, and fit to the data).

Within-person variability can be modeled with various predictors besides time. We chose, however, to specifically highlight the importance of time as a factor because of the surprising frequency with which it has been overlooked in diary research. One of the greatest strengths of diary designs is their ability to characterize temporal dynamics, such as diurnal cycles, weekday versus weekend effects, seasonal variation, or the effect of time to, or since, an event. We urge researchers to use diary designs for this purpose and to investigate temporal effects in existing diary data.

MODELING WITHIN-PERSON PROCESSES: WHAT IS THE WITHIN-PERSON PROCESS FOR THE TYPICAL PERSON, AND HOW DO PEOPLE DIFFER IN THESE PROCESSES? The most challenging questions that can be studied using diary designs are those that address the processes that underlie within-person variability. Diaries can help determine the antecedents, correlates, and consequences of daily experiences. They

can also be used to evaluate whether individuals differ in these processes, and if so, determine the sources of these individual differences.

Many diary studies have focused on questions of this sort, characterizing within-person processes and individual differences. For example, Almeida et al. (2001) investigated influences on the quality of father-child interactions. Mendoza-Denton et al. (2002) investigated the effects of race-based rejection experiences on adjustment to college. Along the same lines, a possible question in our ongoing example would be the identification of factors affecting intimacy on a moment-by-moment basis. Consider an investigation of the influence of social support and social hindrance on feelings of intimacy among married women (cf. Vinokur et al. 1996). The researcher may request participants to complete diary entries following instances of perceiving support or hindrance from their partner. The relevant instances could be of any duration and may be appropriate for self-report so long as they fit the investigator's preestablished definition. Diary data have been used in a similar application to examine the link between perceived partner responsiveness and intimacy across a range of social interactions (Laurenceau et al. 1998).

In such a study, the researcher must also decide on the rate and timing of self-reports. If these are high-frequency events, he or she may wish to ask participants to report only on some instances or on a certain number of instances each day [cf. work done with the Rochester Interaction Record (RIR) (e.g., Wheeler & Nezlek 1977, Tidwell et al. 1996)]. Allowing this choice runs the risk of introducing each participant's bias in selecting some instances and overlooking others. This risk can be minimized with appropriate subject training (such as that conducted in RIR studies).

The researcher must also choose whether to have participants report on the relevant events immediately after their occurrence or to allow participants to postpone responding at inopportune moments (recognizing that the latter concession may work against the recall accuracy, or reliability, of the reports). The researcher may request completion of entries at the conclusion of appropriate events, with the stipulation that they may be postponed under specified circumstances.

Diary studies of within-person processes are mostly nonexperimental, allowing putative causes and consequences to occur naturally. Like all nonexperimental designs, they at best allow weak inferences about cause and effect. However, as we will discuss in more detail below, diary designs are superior to traditional designs in examining processes because they allow investigators to examine the temporal sequencing of events and to control for third variables by using participants as their own controls. To do so, diaries are commonly analyzed using methods of longitudinal data analysis. For example, in a study of the effects of receiving support on participants' mood, Bolger et al. (2000) controlled for participants' previous day's mood. Kennedy et al. (2002) examined conflict and anger at the end of the day while adjusting for anger reported upon waking that morning.

Finally, few diary studies have capitalized on major events and transitions to study psychological change and to model the processes underlying it. Examples are school and family transitions, and scheduled health and occupational events.

Such studies are powerful because they target periods when people and their environments are in flux, and they yield fine-grained data on mediating variables during such periods. For recent examples, see Bolger et al. (2000) and Mendoza et al. (2002).

SUMMARY Diary research is most effective when the design and the research question are complementary in form. As the previous three sections illustrate, well-designed diary studies can answer questions regarding aggregates of experiences over time, temporal patterns of experiences, and the factors affecting changes in these experiences. For each type of question, diary studies can provide information about the average person, between-person variability, and predictors and determinants of this variability. Table 1 provides examples of such questions.

Types of Diary Designs

After identifying the questions that guide an investigation, researchers are faced with concrete decisions regarding design, technology, and data analysis. Diary studies have often been classified into the three categories of interval-, signal-, and event-contingent protocols (e.g., Wheeler & Reis 1991). The interval-contingent design, the oldest method of daily event recording, requires participants to report on their experiences at regular, predetermined intervals. Signal-contingent designs rely on some signaling device to prompt participants to provide diary reports at fixed, random, or a combination of fixed and random intervals. Event-contingent studies, arguably the most distinct design strategy, require participants to provide a self-report each time the event in question occurs. This design enables the assessment of rare or specialized occurrences that would not necessarily be captured by fixed or random interval assessments.

As we see it, diary studies serve one of two major purposes: the investigation of phenomena as they unfold over time, or the focused examination of specific, and often rare, phenomena. It appears to us that the three-way classification blends this conceptual distinction with the technological issue of signaling. Instead, we incorporate interval- and signal-contingent designs into a single category, which we call time-based designs.

TIME-BASED DESIGNS Most diary studies, particularly those focused on within-person processes, involve time-based designs. Examples of these include single-day units [e.g., exploring end-of-day intimacy in relationships (e.g., Laurenceau et al. 2002) and daily levels of stress and mood (e.g., Bolger et al. 1989)] and hourly variations in specific processes or experiences [e.g., examining time-of-day effects on felt-intimacy levels, or within-day fluctuations of moods (e.g., Rafaeli & Revelle 2002)]. Additionally, questions about the aggregate levels of a variable (for example, mean levels of intimacy) are often most appropriately studied using time-based designs, which can be used to systematically sample participants' responses within a day, week, or year. Research with time-based design is often concerned

with ongoing experiences that can be assessed within the course of a typical period. Rare or highly specified experiences are difficult to assess using a time-based design and should be investigated using the event-based design, discussed next.

Investigators designing time-based studies must decide on the most suitable interval for assessment. Assessment times can follow fixed, random, or a combination of intervals. In fixed-schedule designs, researchers choose schedules that include either specific times of the day (e.g., 10 AM, 4 PM, 9 PM) or specific time intervals (e.g., every three waking hours, every evening). Variable schedules involve the assessment of experiences according to a random pattern that is undisclosed to the participants. Users of variable schedules are often concerned with momentary experiences such as psychological states, as in the case of the classic work on adolescent self-esteem by Larson & Csikszentmihalyi (1978).

The selection of schedules and intervals should be theoretically and/or empirically guided. Some phenomena show considerable regularity and can be studied using fixed schedules. For example, a researcher interested in dietary habits may find a fixed-time schedule of morning, afternoon, and evening reports to be most appropriate because it captures individuals' traditional eating behaviors. Phenomena such as intimacy, mood, or other ongoing experiences may also be best addressed with a fixed-interval schedule (e.g., every three waking hours, once per day). A variable schedule may be most appropriate for other phenomena, particularly those sensitive to reactivity or expectation.

Fixed schedules When using a fixed-time schedule, the spacing of intervals is an important consideration. Too long an interval may obscure natural cycles (e.g., diurnal rhythms) or exclude important intervening events and processes. Lengthy intervals may also contribute to the risk of biased retrospection which diary research explicitly seeks to minimize. More distant events are less likely to be recalled accurately; they are also more likely to be influenced by retrospection and by current psychological states (cf. Shiffman et al. 1997).

Retrospection bias may be a more pronounced problem for some phenomena than for others. Concrete, objective events (e.g., number of caffeinated beverages consumed) may be less susceptible to recall bias than are transient subjective feelings such as pain or mood. For example, retrospective reports of pain experiences may be highly influenced by the general level of pain experienced, making it difficult to accurately report the absolute level of the pain experienced each day (Redelmeier & Kahneman 1996).

Intervals that are too short may also be problematic. One possible problem is an unfavorable signal-to-noise ratio. Specifically, researchers may miss slower-acting processes (e.g., day-to-day, week-to-week changes) if data are collected and analyzed at intervals that are much shorter than what is needed to capture the change process (e.g., hourly). Another possible problem is the considerable (and perhaps unnecessary) burden placed on participants by frequent reporting.

Many areas of psychology have yet to develop theories that specify the appropriate time lag for a given phenomenon (see Gollob & Reichardt 1987). Nonetheless,

when diary researchers elect to use fixed-interval response schedules, they need to decide at the outset what interval is most appropriate. Two questions are paramount: What time lag is acceptable between the experience and the description, and what time frame is likely to reveal dynamic processes that are of interest? As a rule, it is better to err on the side of shorter intervals, because the data can then be analyzed with lags of different lengths.

Finally, while a fixed time interval has the benefit of being well suited for longitudinal data analysis, we caution researchers against selecting this (or any other) design feature solely on the basis of data analytic concerns. Such choices carry the risk of producing findings that do not meet the particular operational circumstances and theoretical goals of the research. As we discuss in the section on Diary Data Analysis, below, we believe analysis and design decisions should be considered together.

Variable schedules For certain purposes, researchers may choose variable- or combination-(variable plus fixed) schedule designs. The major benefit of these is their ability to randomly sample moments in a participant's day. Take, for example, a research project investigating the frequency of stress experiences among students approaching an exam. Assessing the participants at random, and inquiring about the present moment, may reduce the potential for biased reports. In contrast, a fixed schedule may sensitize participants to stress because of the routine anticipation of reporting times.

A potential limitation of the random as opposed to fixed schedule is participant burden. Diaries become intrusive when participants are required to complete entries immediately following randomly timed signals. With fixed schedules, the burden is smaller because the timing of diary entries is predictable and can (to some extent at least) be accommodated into a participant's schedule. To ease the intrusiveness of randomly timed signals, researchers may allow participants to postpone responding at inopportune moments. However, allowing delays in the completion of an entry may introduce bias, whereby participants select the times on which they want to report.

Both fixed and random time-based designs allow (and at times require) the use of signaling or notification. Participants can describe their current activity when signals are delivered; alternatively, a researcher may decide to direct participants in advance to follow a certain response schedule. We discuss some practical considerations of signaling in the section on Technology, below.

EVENT-BASED DESIGNS Event-based designs require participants to provide reports at every instance that meets the researcher's preestablished definition. An event-based design is usually most appropriate for diary studies of specific classes of phenomena or processes, especially those that are isolated and/or rare. For example, Jensen-Campbell & Graziano (2000) reported on an event-based study of interpersonal conflict using the RIR in early adolescence; similarly, Laurenceau et al. (1998) used this design in examining the link among disclosures, responsiveness,

and intimacy. In our running example, a study of the association of feelings of intimacy with occurrences of support and hindrance illustrates a typical use of event-based diaries.

The event-based design requires a clear definition of the triggering event(s). Any ambiguity as to which events fall within that definition may lead participants to omit relevant exemplars. To reduce ambiguity, those designing an event-based diary study should consider identifying a single class of events as focal (e.g., Laurenceau et al. 1998). Examining multiple classes of events greatly increases the possibility of confusion as to whether a given event should be reported, as well as the risk of participant burden. Both these risks may lead to a decrease in the number of episodes reported and a weakening of the study's usefulness.

The focus of event-based studies on selected events is both their strength and their drawback. This selection carries at least two risks. The first risk is that participants may not reliably identify each relevant event. The second risk lies in overgeneralizing from the event-based responses to the person's general experience. This risk may exist even when all relevant events are comprehensively reported; for example, the hypothetical researcher examining feelings of intimacy in response to support and hindrance situations may mistakenly draw conclusions about average or typical intimacy levels, when in fact the reports about intimacy were all generated at atypical times.

Time- and event-based designs are not mutually exclusive. Mixed or combination schedules can markedly strengthen a study design. For example, Mohr et al. (2001) combined time- and event-based methods to investigate interpersonal experiences and alcohol consumption in different contexts. Interpersonal events were recalled at end of day with time-based (i.e., daily) diaries, whereas drinking behaviors were monitored with event-based diaries completed throughout the day. In another possible combination design, the occurrence of a particular event (e.g., an argument, a stressful event) triggers a series of subsequent diary reports that trace the temporal sequelae of the event (e.g., Stone et al. 1998).

SUMMARY Diary studies typically focus either on unfolding dynamic phenomena or on specific (and often rare) events. The appropriate design for a study follows directly from its focus, and each design has some benefits and some liabilities.

Limitations in the Use of Diaries

Although diary methods offer many benefits, especially when compared to traditional survey designs, it is important to consider what is known about their problems and limitations. One practical concern is that diary studies often require detailed training sessions to ensure that participants fully understand the protocol (Reis & Gable 2000). Moreover, in order to obtain reliable and valid data, diary studies must achieve a level of participant commitment and dedication rarely required

in other types of research studies. The burden of repeated queries and responses places substantial demands on the participant. To address this, investigators usually design diary instruments that are short and take several minutes to complete. Doing so can limit diary studies to less in-depth reporting of a phenomenon at each time of measurement.

Little is known about the effect of diary completion itself on participants' experience or responses. Several effects, including reactance, habituation, increased complexity, and gradual entrainment, are possible, particularly in more socially reactive behaviors. Reactance refers to a change in participants' experience or behavior as a result of participation in the study. At present, there is little evidence that reactance poses a threat to diaries' validity. For example, Litt et al. (1998) reported that although their participants noted being more aware of the monitored behavior, the behavior itself was not reactive. In several diary studies, Gleason et al. (2001) have documented negative mood elevation in the initial days; in each, the initial spike in negative affect was short-lived, and it dissipated within two to three days. These authors argue that diaries may lead to less reactivity than other forms of data collection because of a habituation process. On the other hand, habituation, and more specifically the development of a habitual response style when making diary entries, may have some deleterious effects. For example, participants may develop a tendency to skim over sections of a diary questionnaire that rarely applies to their experience and may omit responses even at relevant times.

With repeated exposure to a diary questionnaire, participants' understanding of a particular construct may change in at least three possible ways. First, a more complex understanding of the surveyed domain may develop, as may enhanced encoding or retrieval of domain-relevant information. No study has directly tested this, but Thomas & Diener (1990) provide some indirect evidence against this possibility, at least with mood: Accuracy in recalling moods did not differ following an intensive diary period. Second, while participants may not develop a more elaborate or complex knowledge of the monitored domain, the experience of the diary study may entrain their conceptualization of the domain to fit with those measured in the diary. For example, a study of daily intimacy inquiring about three types of intimacy indicators may steer participants to become conscious of these indicators and to be less sensitized to others. Third, research by Pennebaker and colleagues (e.g., Suedfeld & Pennebaker 1997) has documented therapeutic outcomes to a certain kind of self-reflective recollection process. These outcomes, however, have not been found in simple quantitative ratings such as are typical in most diary studies.

Finally, not enough is known about personality (e.g., conscientiousness) or symptom [e.g., alcohol consumption (cf. Litt et al. 1998)] effects on response compliance or styles. Additionally, individual differences in various dimensions (e.g., learning disabilities) may lead to selective biases in the ability to respond to diary questionnaires. We concur with Reis & Gable (2000), who suggest that these and other potential effects of diary methods should be investigated both for methodological and theoretical reasons.

Design Section Summary

We urge (a) care in selecting a design that captures the temporal patterns (if a time-based design is called for) or typical base rate (if an event-based design is called for) of the phenomenon under investigation; (b) minimization of the limitations inherent in diary design, such as participant burden and response bias, by understanding their origin; and (c) recognition of the variety of questions that can be asked about the phenomenon (examples are detailed in Table 1).

DIARY STUDY TECHNOLOGY

This section describes the evolution of diary research technology. After close to 40 years of relying on simple paper and pencil methods, two major waves of change in diary research occurred in the last two decades: the augmentation of paper diaries with signaling devices and the emergence of electronic means of data collection. The most recent technologies promise to increase the ease of providing diary reports and to integrate these reports with physiological and other collateral measures.

Paper and Pencil Diaries

Paper and pencil (P&P) diaries were the earliest and are still the most commonly used approach in diary research. Studies using this approach began appearing in the 1940s (e.g., Stonborough 1942; cf. Allport 1942). Both the technology and the benefits of simple P&P studies are clear. Participants are equipped with folders, booklets, or packets of questionnaires, one for each diary entry. In a pretesting session, they are instructed on how to complete and return the diaries; in some studies, they complete the first diary entry in the laboratory. The process of setting up P&P studies does not differ from most other questionnaire studies. Additionally, because P&P forms or questionnaires are familiar to participants, P&P is the easiest technology for participants to use. Diary studies using simple P&P techniques have several prominent limitations, however, most of which have been highlighted elsewhere (e.g., Feldman Barrett & Barrett 2001, Shiffman & Stone 1998). The simplest one is the risk of honest forgetfulness, where participants fail to remember the scheduled response times (in time-contingent protocols) or fail to have the diaries at hand (in both time- and event-contingent protocols). This is compounded by the risk of retrospection error, where participants rely on (benign) reconstruction or (deliberate) fabrication to complete missed entries, potentially defeating the main benefit of diaries, namely, their ability to obtain accurate, real-time information.

Both honest forgetfulness and retrospection error are related to a third problem, uncertain compliance. Compliance needs to be considered in terms of both the number of entries and their validity. The former is easy to estimate; the latter,

unfortunately, is not. Several researchers have used debriefing interviews to obtain self-reported compliance indices, with mixed results (cf. Feldman Barrett & Barrett 2001). Others have attempted to use verification information; for example, Rafaeli & Revelle (2002) asked participants to report the time of response, while simultaneously obtaining collateral information from a separate computerized task, thus identifying those who are prone to inaccuracy. Though useful, neither of these techniques fully guarantees compliance.

Uncertain compliance is likely to be more problematic in some areas than in others. For example, studies of temporal patterns of responses or of fleeting subjective experiences (such as pain or mood) might be more adversely affected by uncertain compliance. The same is true for studies of phenomena that require detailed encoding of particular episodes (cf. Shiffman 2000).

An additional limitation in P&P diaries is the absence of response-time information, which may be quite valuable. Researchers who are interested not only in the actual response, but also in questions of automaticity, salience, or certainty of response, may find this limiting.

Several of the shortcomings of P&P methods are not specific to diaries and occur in any questionnaire study. However, in the context of some diary studies, these problems may be compounded because of the enormous quantity of data amassed. One of these is human error, either in the response or in the entry stages. For example, it is not uncommon for participants to skip whole sections of a P&P questionnaire; in a multiday diary study, this behavior often goes undetected until the end of the data collection, when it is too late to correct.

Another shortcoming that exists with other self-report methods but increases in scale in diary studies is that of burden of data entry and handling. Even in relatively small diary studies, the data records number in the thousands. To minimize data entry errors it is necessary to double-enter at least a sample of these data.

Finally, because of the risk that prior days' responses in P&P diaries may be viewed by others in their environment, participants may hesitate to be completely truthful in their responses about interpersonal events and emotions. This concern can be mitigated, however, by asking participants to seal the pages of completed diaries (e.g., Bolger et al. 1989, Laurenceau et al. 2002).

We have found the following suggestions useful in implementing P&P diary studies. First, make the diaries easily portable: Arrange the diaries as pocket-sized booklets, stapled or bound, thus allowing participants to carry them around in one piece. Second, reduce the possibility of participant error: In fixed-schedule studies, for example, preprint the dates and times of expected responses onto the diary sheets in order to keep participants on track. Reduce demand characteristics by clarifying the importance of accurate over numerous responses. Asking participants to note whether the entry was completed "on time" (in a way that recognizes that at times it will not be) can also be helpful; so too can collection of responses at several interim points during the length of the study.

Third, pilot test your diary on participants from the population to be studied. Diary forms that seem straightforward to researchers may pose unforeseen

difficulties for participants. Finally, maintain ongoing contact with participants, in a personal yet nonintrusive manner. Diary studies are time-consuming, and this personal contact retains participants more than do monetary incentives or dependence upon goodwill towards science. This last suggestion applies to diary studies of all methods, paper-and-pencil or otherwise.

Augmented Paper Diaries

Over the past three decades, researchers have developed methods that begin to address the limitations of simple P&P diaries. Historically, these methods have been identified with the experience-sampling method (ESM) (Csikszentmihalyi & Larson 1992) and the ecological momentary assessment (EMA) method [(Stone & Shiffman 1994); note that EMA studies are not restricted to those using signals]. Participants' responses are still collected using P&P questionnaires, but these are augmented with signaling devices, such as pagers (e.g., Dabbs et al. 1997, Larson & Csikszentmihalyi 1978), preprogrammed wristwatches (e.g., Litt et al. 1998), or phone calls (e.g., Morrison et al. 1999).

These augmentations offer a remedy to the first problem of P&P protocols (honest forgetfulness) and some relief for the next two (retrospection and uncertain compliance). Rather than relying on a participant's timeliness or individually devised methods of self-reminder, ancillary devices can be preprogrammed to signal randomly or at fixed intervals, prompting research participants' responses and relieving them of the need to keep track of the appropriate occasions for response. Such methods offer an improvement over simple P&P for time-based protocols but are of limited utility for event-based studies. It may be possible to randomly signal participants and obtain sufficient responses for some high-frequency or prolonged events [e.g., being alone (Larson & Csikszentmihalyi 1978)]. However, time-based signals (whether random or fixed in schedule) are likely to miss many discrete events, even those that occur numerous times each day (e.g., smoking).

The augmented approach keeps the benefits of simpler P&P protocols, particularly from the participant's perspective. The ease of making P&P responses is identical, while the signaling device eliminates the burden of remembering to complete diary entries. At the same time, some of the limitations of simpler P&P designs (retrospection error, uncertain compliance, cumbersome data entry and data management) remain.

From the researcher's perspective, several other factors may reduce the appeal of these methods. First, setup is somewhat more involved and costly. Signaling devices need to be purchased, programmed, and maintained, and participants need to be trained in their usage. Additionally, these methods magnify the problem of disruptiveness. Being beeped to complete a smoking-cessation or relationship-quality diary entry in the middle of a job interview or a first date may be an unacceptable disruption. Potential participants may anticipate this and decline participation or adopt behaviors that defeat the purpose of the studies (e.g., not carrying the signaling device).

To address these problems, researchers need to be aware of their participants' schedules and must choose and program the signals in a way that is minimally

disruptive of the participants' routines (e.g., sleep, classes, meetings, etc.). Though feasible, this often requires added attention, such as tailoring the signaling schedule individually for each participant.

Handheld and Electronic Data Collection

Electronic data collection methods of various forms began appearing over the last decade (Feldman Barrett & Barrett 2001; Shiffman 2000). Typically, these studies use handheld computers (i.e., palmtop computers, personal digital assistants) equipped with custom-designed questionnaire programs.

Several such programs have been developed. Perez & Reicherts (1996) describe one of the earliest programs, COMRES, which has been used fruitfully, especially in Europe. The first to be used extensively was the Electronic Diary (see Stone et al. 1998), which has been applied to various health and social psychology questions for over a decade now (e.g., Paty et al. 1992). Several researchers have developed their own diary programs [e.g., PMC-diary for Palm Pilots (Rafaeli & Revelle 1999); ISIS for pocket PC (Rafferty 2001)]. A major contribution to the field was made by Barrett & Feldman-Barrett (2000), who used a National Science Foundation grant to develop a freeware diary program for handheld devices (Experience Sampling Program, or ESP).

The benefits of such programs are immediately visible. First, they allow for signaling. Second, they provide time-stamps (and date-stamps) for responses. When taken together with appropriately designed questions, such as ones that focus on the present moment, these features obviate the problems of forgetfulness and uncertain compliance. In fact, time-stamping provides a direct measure of compliance. By examining the times of the responses, researchers can immediately identify diary entries that were completed on time, within a reasonable delay period after a signal, or at unacceptable times. Additionally, researchers are often interested in ensuring that entries were completed in their entirety within a certain time. Electronic diaries can document or even enforce this. For example, ESP allows researchers to determine the maximal acceptable length of time for each answer, or for the entire questionnaire, in time-contingent protocols; PMC allows researchers to do so in event-based protocols as well. Both programs have defined times (typically, immediately after the completion of one entry) during which further diary entries are disallowed.

A third benefit available in some of the programs is flexibility in the presentation of questions. Some programs allow randomization (e.g., ISIS, ESP) and/or hypertextuality [i.e., the ability to present different questions based on responses (ISIS)] of the presentation of items. These features allow the diaries to be less cumbersome and repetitive. More exciting still is the possibility of the assessment paradigm adapting dynamically to the provided responses. For example, Shiffman et al. (1997) report on a study of smoking behavior in which the content and timing of questions changed in response to the pattern of earlier responses.

Fourth, electronic diaries offer major advantages in terms of data entry, management, and accuracy. Since participants enter their responses directly into the electronic diaries, the processes of transcribing and double-checking the data,

which are costly and error prone, are bypassed. Additionally, the programs typically prevent responses that are out of range, and they minimize the risk of skipped questions (since the items are typically presented in sequence, ending only when the whole entry has been completed).

Fifth, to varying degrees, these programs simplify the process of taking into account the participant's schedule. For example, ESP allows the researcher to set both start and end times for each participant at the beginning of the diary period. These could be obtained by inquiring about each participant's typical schedule. Somewhat more flexibly, PMC allows participants to put the diary to sleep, scheduling it to wake at a personally convenient time; it also allows "snoozing" the diary for variable lengths of time. As Shiffman (2000) notes, the ability to constrain the diary signals to appropriate times allows a more comprehensive sampling of those times.

In selecting or developing an electronic diary program, researchers may find specific capabilities to be particularly important. Two features relate to flexibility in the presentation of questions: an ability to randomize item presentation (both within blocks and in whole blocks) and an ability to display several different response formats (e.g., Likert scales, check boxes, etc.). These features are particularly useful when the diary questions are of different sorts and when researchers want to avoid response sets.

Several features address the usability of diary programs. Does the program allow participants to review and change their answers (i.e., to back up in the diary)? In ongoing research, we have found that participants strongly favor the ability to retrace one or several steps. Does the program "take over" the electronic device, disallowing access to any other program? Often this is useful in constraining the use of a device (e.g., conserving battery power, reducing the risk of participant access to stored answers). However, as the use of handheld devices becomes more prevalent, researchers may wish to allow their programs to run in the background of participants' proprietary devices. In that case, a diary program that takes over the device would be a liability. Another feature to consider is whether a program allows event-based responses, time-based responses, or both within the same study.

Finally, several features can make diary entries quick or slow to complete. One is the number of questions presented on each screen. Most programs, including ESP (but not ISIS), have the mixed blessing of presenting only one question at a time. This may improve the processing of each question, but it may also lengthen the total time needed to complete the diary and therefore increase the burden. In some programs, tapping or pressing the answer advances the diary. In others, particularly ones where multiple questions appear at once, participants are required to tap a special key (e.g., "NEXT") after each completed page. In lengthy diaries, this extra keystroke adds noticeably to diary completion times.

Although handheld devices are the most popular means of electronic data collection, they are not the only ones. Some researchers have turned to Web-based questionnaires or to phone-in protocols (Searles et al. 2000). One benefit of these methods is their ability to secure the responses immediately while maintaining

many of the advantages of handheld data collection. Their main drawback is inconvenience for participants, who complete a diary entry on a phone or a computer connected to the Internet. As mobile Web access for handheld computers becomes common and affordable, the functional differences among handheld, phone, and Web-based programs will likely disappear.

Promising though they are, electronic diary collection methods do have some limitations. The first is the development cost of the appropriate program. At the time the first electronic diary studies were published, these costs were prohibitive for many researchers. However, recent advances have brought electronic data collection within reach for many researchers. Some find it best to develop a new program tailored to their specific needs (e.g., Perrez et al. 2000, Rafaeli & Revelle 1999, Raffety 2001). Others have adapted existing programs (e.g., Mohr et al. 2001). Finally, the arrival of a federally funded freeware program, ESP (Barrett & Feldman Barrett 2000), means that all researchers have access to a basic program suitable for many purposes.

Costs do accrue at other stages of electronic data collection. Electronic devices require maintenance and may break or need replacement. Nonetheless, development and operation costs can be offset, in part or in full, by savings on data entry, management, and storage, as well as by increased data integrity (which can translate into greater reliability, greater power, and lower sample sizes).

Currently, affordable handheld devices leave much to be desired in technical terms. Font size tends to be small, visual contrast is not very sharp, and the battery power of the devices does not always suffice for longer studies. Additionally, these devices are best suited for handling simple, close-ended responses. The incorporation of open-ended responses remains difficult at present. However, given the speed of innovation in the design of handheld devices, these technical limitations can be expected to lessen greatly in coming years.

Other limitations of electronic data collection methods are likely to be more enduring. Key among these are the cost and resources needed to train participants in the use of such devices. Typically, researchers have found it necessary to devote at least 30 minutes, and often more, to acquainting participants with the use of an electronic diary. This training is usually carried out individually and in person. Thus, for large samples, and for studies where many participants never meet face-to-face with the researchers (e.g., Bolger et al. 2000), the use of electronic devices can be problematic.

Additionally, a concern for the field as a whole is the possible risk of perpetuating a "digital divide," i.e., differential access of individuals of various groups to participation in research. For example, researchers may hesitate to use high-cost devices in studies of poverty. Such hesitation could subtly steer research towards more convenient or accessible populations and away from economically disadvantaged participants. Similarly, computer literacy may affect participants' willingness to enroll in electronic diary studies. Clearly we need more published studies on the feasibility of using electronic diary studies in a broader range of special populations.

Emerging Opportunities

We and others (e.g., Shiffman 2000, Feldman Barrett & Barrett 2001, Feldman Barrett & Jamner 2000) see several avenues of opportunity in the technology of diary studies. These can be divided broadly into two types: continued improvement in self-report diary technology, and integration of collateral information collection into research protocols.

IMPROVING SELF-REPORT TECHNOLOGY At least two types of advances in self-report methodology are in progress. First, improved mobile communication allows online, duplex (i.e., interactive) contact with participants. For example, researchers can now send questions to, and receive responses from, participants in real time. This allows a dynamic tailoring of questions for specific participants; at the very least, it offers an additional method for increasing compliance by maintaining contact with participants (e.g., Jamison et al. 2001).

Second, advances in voice recording and recognition as well as in linguistic analysis technology allow the inclusion of verbal reports into diary research. Rather than relying only on close-ended questions, researchers are now able to request open-ended responses, or to sample speech naturalistically, examining these responses within the context of daily functioning (e.g., Pennebaker & Graybeal 2001).

INTEGRATION OF COLLATERAL INFORMATION The inclusion of information that goes beyond self-reports offers much promise for a more comprehensively ecological psychology. The technology allowing such integration is already available and in use, particularly in the field of behavioral medicine. For example, Jamner et al. (1998) describe research examining the concordance of heart rate, blood pressure, and subjective mood that incorporated ambulatory monitors of the physiological indices. Shapiro & Goldstein (1998) used similar physiological indices along with an objective monitor of movement, the wrist actigraph, which allows accurate tracking of physical movement during both wakefulness and sleep. Multichannel recorders are now capable of jointly tracking indices of heart rate, blood pressure, respiration, and physical activity; it is only a question of time before unified devices allow an integration of these indices with self-reported written or spoken responses. (For a review of several ambulatory monitoring techniques in psychophysiology and psychology, see Fahrenberg & Myrtek 1996.)

Technology Section Summary

Innovations in diary research technology are rapidly occurring. Various researchers are working to improve our technical ability to conduct ecologically valid research that would be minimally intrusive and maximally reflective of individuals' ongoing feelings, thoughts, goals, behaviors, and circumstances.

DIARY DATA ANALYSIS

Data analyses of diaries address the three types of research questions discussed earlier: What is an individual's typical or average experience, and how do people differ in these averages; what is the time course of a person's experiences, and how do people differ in their time courses; and what processes underlie changes in a person's experiences, and how do people differ in these processes? We discuss general statistical approaches to each type of question; where necessary, we also discuss special data-analytic issues that arise in certain types of diary studies (e.g., estimating measurement models, analyzing dyadic diary data).

Statistical methods for diary data analysis must address several key issues that make these data unique. First, the data involve repeated measurements, and thus the within-person data points cannot be assumed to be independent. Second, the dependence of the observations is often serial, in that diary reports adjacent in time are more similar than those more distant in time. Third, the number of diary reports is often large and varies from person to person, making usual repeated-measures analyses infeasible. Finally, temporal patterns and cycles are often present in the data, and thus flexible classes of mathematical models need to be considered. A consensus is emerging among diary researchers that multilevel models (also called hierarchical linear models, general mixed models, and random regression models), modified to handle repeated-measures data, are appropriate for diary data analysis. Although these models are complex, we believe they are in important respects easier to intuit than other complex, widely used models such as structural equation modeling (SEM) or meta-analysis.

Those wishing instruction on multilevel models can consult major textbooks geared toward social scientists (Hox 2002, Raudenbush & Bryk 2002, Snijders & Bosker 1999). A brief introduction to multilevel models is provided by Kenny et al. (1998). Reis & Gable (2000) also discuss the benefits of multilevel analysis in their review of diary work in personality and social psychology. Also useful are two edited volumes on longitudinal data analysis (Moskowitz & Hershberger 2002, Sayer & Collins 2001).

A number of sound and flexible multilevel modeling programs are currently available. The most commonly used in diary research are two freestanding programs [HLM (Bryk et al. 1996) and MLWin (Prosser et al. 1996)] and one module of the SAS statistical software system called PROC MIXED (SAS Institute 1999). For readers interested in working with PROC MIXED, Singer provides an especially useful introduction to basic analysis issues for estimating time-course models with longitudinal data (Singer 1998, 2002). Finally, note that specific multilevel models [i.e., for growth curve analysis (Willett & Sayer 1994)] can be estimated within an SEM framework and that more general multilevel capabilities are beginning to be available in multivariate software systems such as LISREL (Jöreskog & Sörbon 1996), EQS (Bentler 1995), and Mplus (Muthén & Muthén 2002).

Aggregating Over Time: What is the Typical Person Like, and How Much Do People Differ from Each Other?

The first approach uses only averaged, summary data on the repeated measurements obtained from the diary. Although diary designs by necessity obtain repeated measurements on participants, researchers may not necessarily be interested in temporal or within-subject patterns in the data. Rather, the researcher may choose simply to summarize each person's within-person data in the form of averages (e.g., means) and variability (e.g., variances) and to examine between-person averages and variability in these summary measures.

An example of this approach is a study where a researcher uses a time-based weekly diary design to document intimacy levels in married women over a 16-week period. We can expect such a design to result in a more valid picture of daily intimacy over that period than would be obtained from a retrospective report summarizing the entire four months.

THE TYPICAL PERSON'S AVERAGE AND VARIABILITY Figure 1 shows a graph of hypothetical data for a sample of four women over the 16 weeks. Since the first set of research questions ignores the patterning over time, we begin by examining the small bell-shaped curves along the Y axis of the graph. These are summary distributions of weekly intimacy scores for each of the four women. Taking participant

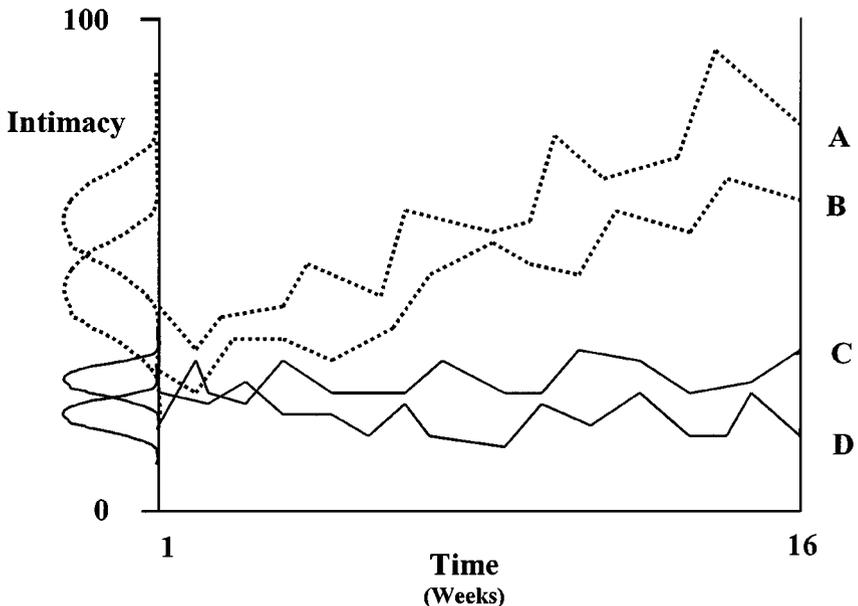


Figure 1 Hypothetical data of four participants in Intervention Study (16-week weekly diary study of intimacy following brief intervention or control).

C as an example, we can see (a) how intimate she felt with her partner on a typical week over the 16-week period (about 35 units on a 0–100 scale); and (b) to what extent her level of intimacy varied over time (the participant's standard deviation is about 3 units).

In this example, there are visible differences between women in their typical levels of intimacy and in their variability in intimacy, with those with lower levels of intimacy showing less variability. Nonetheless, we may be interested in characterizing the mean and variance of an average person in the sample. To do so we can calculate (a) the average level of intimacy for the average person (about 40 units) and (b) the variability of intimacy for the average person (about 4 standard deviation units).

Estimates of the mean and standard deviation of intimacy for the typical woman can be obtained in two main ways. The first is simply to calculate a mean and standard deviation separately for each participant and obtain arithmetic averages of these to estimate the mean and standard deviation for the typical participant. This approach will suffice if each participant provides many repeated measurements (Kenny et al. 1998). However, a better approach, and one that will pay particular dividends when research questions become more complex, is to use a multilevel analysis.

The multilevel analysis has us specify a statistical model with two levels: level 1, a within-person level, and level 2, a between-person level. The within-person level specifies that each woman's intimacy score is composed of her average intimacy plus that score's deviation from her average. Similarly, the between-person model specifies that a woman's mean intimacy score is composed of the overall mean score across all women and that particular woman's deviation from the overall mean. For more statistical detail on this model, see Raudenbush & Bryk (2002, ch. 2), where it is called a One-Way ANOVA with Random Effects. Using this multilevel model, one can obtain a summary measure of variability for the typical person: the level-1 error variance.

As noted earlier, for the data illustrated along the Y axis in Figure 1, the average woman had an average intimacy of approximately 40 units, and the variability around her average was approximately 4 units. One must be careful when calculating these averages, however. Although in many diary studies the notion of an average person is useful, in our particular example it is misleading because there are two distinct subgroups.

ASSUMPTIONS ABOUT WITHIN-PERSON VARIATION Diary data are longitudinal. Even if the investigator has no direct interest in time as a factor, the data are ordered in time and this ordering may be relevant to one's analyses. In this respect, diary data are often different from other nested data. For example, when the nesting involves persons within groups, the ordering is often inconsequential. In contrast, with the intimacy data in our example it is plausible that adjacent diary reports are more similar than reports farther apart. This will occur if time bears any systematic relation to scores on intimacy (e.g., if there is growth or decline

in intimacy over time). It can also occur even in the absence of a systematic effect of time, in cases where unmeasured influences on intimacy (e.g., child-care problems) have some stability from day to day, thereby leading the daily reports to be dependent due to this common influence. When adjacent data have stronger dependence on one another than nonadjacent data, a pattern of autocorrelation will be observed (West & Hepworth 1991).

Because the correlation between adjacent data points is typically positive (when it exists), autocorrelation generally leads to estimates of within-person variability that are smaller than they would be if the reports were independent of one another. This produces a downward bias in standard errors and overly liberal tests of significance. Available multilevel software such as HLM and PROC MIXED in SAS allow one to model autocorrelation in within-person error terms and correct for this biasing influence. As we will see, though, autocorrelated errors in simple models such as the one we have discussed may disappear when these models are expanded to include omitted variables, such as the child-care problems cited in the example above (Singer 2002). Introductory discussions of autocorrelation in multilevel models can be found in Hox (2002) and Snijders & Bosker (1999). Advanced treatments can be found in Diggle & Liang (2001), McCulloch & Searle (2001), and Verbeke & Molenberghs (2000).

BETWEEN-PERSON DIFFERENCES IN AGGREGATED DATA The extent to which there are between-person differences in a level-1 variable (intimacy in our example) must be determined in models of between-person differences. As drawn in Figure 1, the summary distributions clearly indicate that women differ in their average intimacy over time. With actual diary data there is often a large degree of overlap in individuals' distributions, and appropriate statistical tests are needed to determine if these distributions differ from one another. Further emphasizing the need for appropriate statistics, it is known that distributions of sample means such as these will show more variability than would true means (Raudenbush & Bryk 2002). This is because the observed mean of each woman's intimacy will be an imperfect estimate of her true mean due to sampling error (and possibly measurement error). Not surprisingly, the uncertainty in the estimate of between-person variability is greatest when within-person sample sizes are small.

Fortunately, a multilevel analysis, specifically the One-Way ANOVA with Random Effects described earlier, will produce unbiased estimates of the between-person variability. Recall that level 1 of this model specifies that each woman's intimacy score is composed of her mean intimacy plus that score's deviation from the mean. Similarly, the level-2 model specifies that a woman's mean intimacy score is composed of the overall mean score across all women, and that particular woman's deviation from the overall mean.

All multilevel modeling software can handle models of this kind and all provide (a) a significance test of whether the data show between-person variability in means and (b) an estimate of the size of that variability (as a variance or standard deviation). Likewise, these programs can allow for (and provide tests of) individual

differences in the within-person variability, which, of course, are a feature of the example data. For more detail on this specific model, see Raudenbush & Bryk (2002).

EXPLAINING BETWEEN-PERSON DIFFERENCES To the extent that people do differ from one another, the notion of a “typical person” is less useful in describing the data and suggests the need to investigate sources of between-person variability. In nonexperimental diary studies these can involve relatively stable characteristics of participants (e.g., personality) or of their environment (e.g., the quality of the neighborhood they live in). In experimental work (e.g., when testing an intervention program), one may be interested in comparing the levels of the outcome in an experimental group and in a control group. Assume that the data presented along the Y axis of the graph in Figure 1 are the result of an intervention to boost intimacy in a sample of women from distressed couples. Further assume that the two women with the highest mean intimacy score (A and B, represented by dashed distributions) are from the intervention group and the remaining two (C and D, represented by solid distributions) are from the control group.

The effect of the intervention on average intimacy can be analyzed using a means-as-outcomes model (Raudenbush & Bryk 2002). Level 1, the within-person level of the model, as before, specifies that each woman’s intimacy score on a given week is the sum of her average across all weeks plus that week’s deviation from her average. Level 2, the between-person level of the model, now specifies that each woman’s average intimacy is the sum of the mean intimacy for women in her intervention group (experimental versus control) plus her deviation from that mean. Because Figure 1 shows that the within-person variability differs by condition, the multilevel analysis would need to allow for this to obtain optimal tests of significance. A test of autocorrelation in the within-person errors is also advisable, and if present, the autocorrelation should be modeled, thereby obtaining unbiased estimates of within-person variances.

Modeling the Time Course: How Does a Typical Person Change Over Time, and How Do People Differ in Change Over Time?

In addition to the aggregate analyses described in the previous section, time-based diaries allow investigators to examine changes in levels of variables over time. Using our intimacy example, researchers might be interested in following the women over time to examine changes in intimacy. When considering the time course of intimacy for the average woman, two basic questions can be asked: (a) What was an average woman’s intimacy level at the beginning of the diary study, and (b) what was the average woman’s time course in intimacy?

THE TYPICAL PERSON’S TIME COURSE The data shown in the body of Figure 1 reveal the time course of intimacy for the four women. We see that the four women’s

intimacy levels are very similar at time 1, that there is a divergence over time, and that the growth process for each woman is approximately linear. Even though in this example women differ in their time courses, we can still average over these differences and characterize the intimacy of someone who is average in terms of starting values and rates of change.

An appropriate modeling approach here would be a linear growth model. We specify at level 1 or the within-person level that a woman's intimacy score on a given week is the sum of (a) her starting level; (b) her rate of change per week, multiplied by the number of elapsed weeks; and (c) a residual term, her actual score's deviation from her predicted value for that week. At level 2 or the between-person level, we specify that each woman's starting value (or rate of change) is the sum of the mean starting point (or mean rate of change) plus her deviations from those averages. From this model, the starting value, rate of change, and residual variance for the typical person can be obtained.

BETWEEN-PERSON DIFFERENCES IN CHANGE In addition to estimates for the typical person, the multilevel model just described will provide estimates (and tests of significance) of between-person variability of the starting values and rates of change in intimacy. Note again that the observed variability in starting values and rates of change, if these were obtained from separate regression analyses instead of a multilevel model, would be biased upwards. In the case of slopes based on small numbers of observations, this bias can be severe (Raudenbush & Bryk 2002).

Figure 1 was constructed so that after allowing for individual differences in starting values and rates of change, the residual variances of the growth model at level 1 would not differ across women (whereas the within-person residual variances around the simple mean in the previous section did differ). Other things being equal, large individual differences in slopes will result in individual differences in residual variances in *Y* at level 1 unless the cause of these differences is accounted for in the model. Also, allowing for individual differences in growth can eliminate autocorrelation in within-subject errors, as demonstrated by Singer (1998, 2002).

EXPLAINING BETWEEN-PERSON DIFFERENCES IN CHANGE The data in Figure 1 were drawn to illustrate the time courses of intimacy for women in an intervention study. In the example, random assignment was effective: The two groups did not differ in their initial levels of intimacy. The control group shows essentially no change over time whereas the treatment group shows an increase, on average. The multilevel model that can be used to explain between-person differences is a modification of the one used in the previous section. In this case, the level-1 model remains the same, but the level-2 or between-person model provides estimates of a mean intercept and a mean slope for each condition. Within each condition, individual women are allowed to show deviations from the average for their condition. For example, although the experimental group shows increases in intimacy over time, one woman (A) in the group has a slightly higher rate than the other (B).

Modeling Within-Person Processes: What is the Within-Person Process for the Typical Person, and How Do People Differ in These Processes?

The third and most complete approach to understanding diary data involves developing an explanatory model of the factors affecting within-person variability in the data. Such models may evolve from an initial model of the time course in the dependent variable (e.g., Bolger & Eckenrode 1991) or from an interest in explanatory variables that are not necessarily related to elapsed time in the study (e.g., Swim et al. 2001). Whether or not temporal changes are of interest in one's study, based on our experience it is wise to include in process models parameters for elapsed time in the study and for cyclical effects such as time of day or day of week. Some factors of interest (e.g., social support) may show considerable temporal variability, such as weekday versus weekend differences, and any attempt to estimate support effects should hold constant day of the week, as it is a plausible third variable in the relationship between support and distress.

To illustrate a process model, we continue with a focus on intimacy as a dependent variable but change the example to one of a daily diary study of women in a cohabiting relationship over the course of a typical month. If one wishes to understand the processes underlying day-to-day variability in intimacy, one needs to examine daily events in the relationship that can bring partners together or push them apart. For illustrative purposes, we have chosen to focus on daily conflicts. Daily conflicts have been shown in previous work to be more powerful than any other daily stressor in affecting mood (Bolger et al. 1989).

Figure 2 presents four panels of fictitious data for four women (E, F, G, and H), showing daily changes in intimacy (again on a 100-point scale) over 28 days.

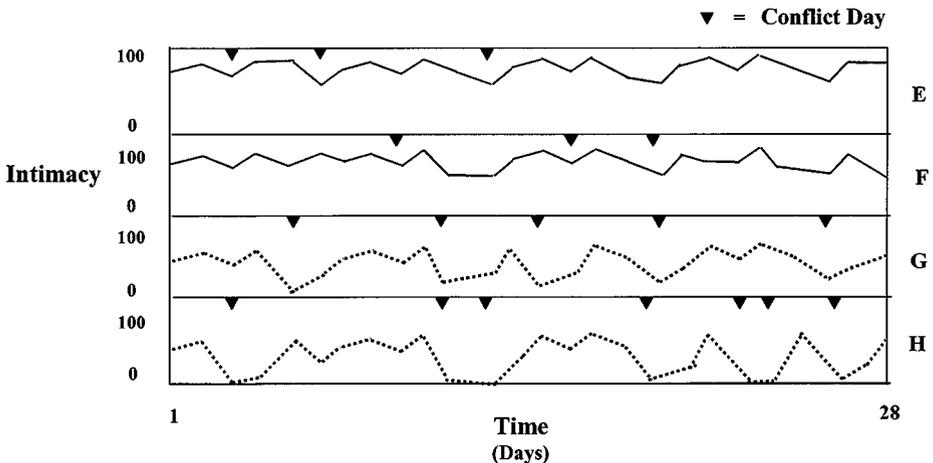


Figure 2 Hypothetical data of four participants in Daily Conflicts Study (28-day diary study of intimacy and daily conflicts).

The downward pointing triangles at the top of each panel indicate days on which a conflictual or hindering interchange occurred between the woman and her partner. Several features of the data are noteworthy. First, unlike the previous example, there are no time trends; for all women gross levels of intimacy do not change over the course of the study. Second, the women show differences in exposure to daily conflicts, with participants E and F showing the greatest exposure and participants G and H showing the least. Third, the women show differences in average levels of intimacy, with participants E and F showing high levels and participants G and H showing moderate to low levels. Fourth, the latter two participants show greater variability in daily intimacy than the former two. Finally, there are marked differences in the extent to which conflicts are associated with declines in intimacy. Participants E and F show small declines whereas participants G and H—particularly H—show much larger declines and ones that last more than a single day.

WITHIN-PERSON PROCESS MODEL FOR THE AVERAGE PERSON AND BETWEEN-PERSON DIFFERENCES IN PROCESS Process models that characterize both the average person and between-person differences present no major new problems in terms of analysis. The main issue to be dealt with is whether, and if so, how, to use the longitudinal nature of the diary design to study lagged effects as a means of addressing the possibility of bidirectional processes. For example, the relation between conflict and intimacy is likely to be bidirectional (low intimacy makes conflict more likely, and conflict leads to decreased intimacy), and it would seem important to demonstrate that earlier conflict predicts later declines in intimacy, controlling for earlier intimacy. This approach is the same as is used in traditional panel analyses, but with a diary design the lagged analysis is within—rather than between—subjects and is therefore less vulnerable to rival hypotheses.

For the example data, a possible approach would be to create two new variables—one-day lagged versions of conflicts and of intimacy—and examine whether, controlling for lagged intimacy, lagged conflict predicts current intimacy. The example data show that such effects are present for participants G and H and absent for participants E and F. One could begin an analysis by estimating separate level-1 regressions for each woman, where the main coefficient of interest would be the one for lagged conflicts. These coefficients could then be averaged to arrive at a between-person average. We have already discussed how the variability in these coefficients would be inflated due to sampling error (and possibly measurement error), but the average coefficient would give an unbiased estimate of the population value.

A more appropriate multilevel modeling approach is what Raudenbush & Bryk (2002) call a Random-Coefficients Regression Model. This model provides appropriate estimates and tests of the average of the level-1 slopes for conflict, and of the variance of these slopes. Such an analysis would show evidence of an average effect of conflict but would also show evidence of between-person differences in this effect. It would likely not show any between-person differences in level-1 residuals. Once the effect of conflict is taken into account, the remaining variability

does not differ across persons (but note that if one carried out a simpler analysis, aggregating the data over repeated measurements, one would find that participants G and H have greater within-person variability than participants E and F).

To keep the exposition simple, Table 1, part III describes a random-coefficients model in which only contemporaneous (cross-sectional) relations between conflicts and intimacy are examined. By replacing contemporaneous conflict with lagged conflict and lagged intimacy, however, a longitudinal analysis can be performed, as discussed above. Level 1 of the model specifies that a given woman's intimacy score on a given day is the sum of (a) her average intimacy score on days when no conflicts occur, (b) the additional decrement in intimacy associated with conflict days, and (c) a residual score, her actual intimacy score's deviation from her predicted value for that day. At level 2 or the between-person level, we specify that each woman's level-1 coefficient is the sum of that coefficient for the average woman, plus her deviation from that average. This model can be used to produce the estimates and tests described earlier: the coefficients for the average woman, between-person differences in these coefficients, and between-person differences in level-1 errors.

EXPLAINING BETWEEN-PERSON DIFFERENCES IN WITHIN-PERSON PROCESSES The example has been constructed to show between-person differences in within-person processes. Assume that participants E and F are women who report high baseline levels of relationship quality (RQ), whereas G and H are those who report low RQ. To investigate the relation between baseline RQ and conflict processes, we now use what Raudenbush & Bryk (2002) call an Intercepts-and-Slopes-as-Outcomes Model. The level-1 model is unchanged, but the level-2 model specifies that between-person differences in intercepts and slopes are a function of RQ. Such an analysis would show that, compared to low-RQ women, high-RQ women have higher intimacy scores on days when no conflicts occur, and show smaller decrements in intimacy when conflict days occur.

Other Data Analysis Issues

MEASUREMENT MODELS FOR WITHIN-PERSON DATA We have just described how theoretical models of causal processes can be estimated using time-based diary data. Readers will have noted that our first and primary concern was to specify the factors that affect within-person variability, how people differ from themselves. An unstated assumption of this approach, however, is that the concepts we wish to model are adequately measured within subjects. With diary data it is rarely clear that this is so. Most validation studies of psychological measures involve between-subject designs, and there is no guarantee that these measures are reliable and valid for assessing within-person variability (Borsboom et al. 2002). Thus diary researchers who adapt well-validated between-subject measures should consider whether the items they propose using have a similar factor structure within subjects similar to that known to obtain between subjects. Furthermore, the

possibility needs to be considered that within-person measurement models differ across people (Molenaar 1999).

To date, there is little published empirical work on measurement models using diary data. Given the importance of this topic, we see it as an area deserving much more research activity. Some notable exceptions include work on the within-subject structure of affective experience (Feldman 1995), and on an integrated measurement and structural model for diary data [using SEM (Kenny & Zautra 1995, 2001)]. A groundbreaking measurement approach that has to date been neglected by diary researchers involves dynamic factor analysis models such as those developed by Molenaar (1985), McArdle & Epstein (1987), and Nesselroade et al. (2002). Future diary work should draw on these and related models when developing and validating measures. One possible reason why measurement issues have been neglected thus far is that measurement models have not yet been incorporated into popular multilevel modeling software. As noted earlier, however, this situation is changing, with programs such as LISREL, EQS, and MPlus incorporating multilevel capabilities. Until technical advances allow measurement models to be routinely included in multilevel models, we recommend that researchers at least confirm the adequacy of measures using within-subject data that are pooled after centering around each subject's mean. This approach will demonstrate that measurement models are sound at the level of the average person for data collected over time. To do so, researchers need to fit a factor model to the pooled within-person data set. An example of such an analysis on 2300 person-days of data on 9 emotion items can be found in Thompson & Bolger (1999).

DATA ANALYSIS OF DIARY DATA ON DYADS, FAMILIES, AND OTHER GROUPS We have argued that diary data can provide a useful window on the processes affecting change in a person over time. Because social interactions are likely to be important determinants of how people change, and because of the importance of close relationships to such interactions, some diary researchers have attempted to study change processes by collecting diary data on dyads, families, and other important social groups (e.g., Almeida & Kessler 1998, Bolger et al. 1989, Larson & Richards 1994, Perez et al. 2000, 2001, Roberts & Levenson 2001).

Although analysis issues become more complicated when diary data on dyads and groups are involved, the statistical models required need not change: Multilevel models can be used effectively for this purpose. Dyadic and group-based diary data can be analyzed in two main ways. The first treats the persons within the larger social group as the main focus and takes account of nonindependence between persons due to their common group membership. In the case of married couples, this involves obtaining separate estimates of within-person parameters for husbands and wives while adjusting for possible correlation in the residual of the model attributable to the couple level of analysis. An important example of this approach can be seen in work by Raudenbush et al. on the relationship between work and family events in married couples (Barnett et al. 1993, Raudenbush et al. 1995). A modification of this approach that incorporates additional sources of

nonindependence—one due to omitted transient couple-level influences (i.e., concurrent within-couple residual correlation), and the other due to omitted transient person-level influences (i.e., day-to-day within-person residual correlation)—can be found in Kennedy et al. (2002).

The second major approach explicitly decomposes the relation between diary measures X and Y into components that reflect covariation operating at the between-group level, between-person within-group level, and within-person level (Kashy & Kenny 2000). This approach is akin to Kenny & la Voie's (1985) work on separating individual and group effects, and to Gonzalez & Griffin's (1999) work on decomposing correlations on data from dyads into dyadic and individual components. A model exposition of this approach can be found in Wilhelm (2001), with analyses from the Fribourg Family Project in Switzerland. The data set consisted of six reports per day for seven days from mothers, fathers, and adolescent children from 96 families. Wilhelm describes a multilevel analysis that decomposes influences on emotional states into between-family, within-family, within-person, and within-day components.

Analysis Section Summary

We have presented an overview of three major approaches to diary data analysis. Often, a single study calls for the use of all three approaches. The classification scheme can be used by researchers to systematically consider the questions they might ask in a diary study, and to determine which ones are interesting given their particular substantive problem. Thus, a researcher may begin by asking, "Do I have any prediction regarding the average (aggregate) level of my variable? What about its variance? Do I expect the average or the variance to vary between people?" and so on, thereby arriving at a data analysis plan.

CONCLUSIONS

Diary research offers a unique window on human phenomenology. As diary researchers, we are clearly enthusiastic about the multiple benefits of using such designs in the study of temporal dynamics, within-person process, and other questions. We hope this enthusiasm proves contagious. At the same time, we have tried to offer a candid look at the dilemmas, irritations, and problems of using diaries to investigate psychological processes. To use diary methods effectively, these problems need to be acknowledged and addressed, but they should not deter researchers from setting out on important studies of the particulars of everyday life.

ACKNOWLEDGMENTS

We are grateful to James Cranford, Lisa Feldman Barrett, Geraldine Downey, Susan Fiske, Marci Gleason, Jean-Phillippe Laurenceau, Meinrad Perrez, Harry Reis, Saul Shiffman, Patrick Shrout, Arthur Stone, Howard Tennen, and members

of NYU's Couples Research Lab for their suggestions during the preparation of this chapter. Work on this paper was supported in part through NIMH grant R01-MH60366 awarded to the senior author.

The *Annual Review of Psychology* is online at <http://psych.annualreviews.org>

LITERATURE CITED

- Affleck G, Zautra A, Tennen H, Armeli S. 1999. Multilevel daily process designs for consulting and clinical psychology: a preface for the perplexed. *J. Consult. Clin. Psychol.* 67:746–54
- Alloy LB, Just N, Panzarella C. 1997. Attributional style, daily life events, and hopelessness depression: subtype validation by prospective variability and specificity of symptoms. *Cogn. Ther. Res.* 21:321–44
- Allport GW. 1942. *The Use of Personal Documents in Psychological Science*. New York: Soc. Sci. Res. Council.
- Almeida DM, Kessler RC. 1998. Everyday stressors and gender differences in daily distress. *J. Personal. Soc. Psychol.* 75:670–80
- Almeida DM, Wethington E, Chandler AL. 1999. Daily transmission of tensions between marital dyads and parent-child dyads. *J. Marriage Fam.* 61:49–61
- Almeida DM, Wethington E, McDonald DA. 2001. Daily variation in paternal engagement and negative mood: implications for emotionally supportive and conflictual interactions. *J. Marriage Fam.* 63:417–29
- Barnett RC, Marshall NL, Raudenbush SW, Brennan RT. 1993. Gender and the relationship between job experiences and psychological distress: a study of dual-earner couples. *J. Personal. Soc. Psychol.* 64:794–806
- Barrett DJ, Feldman Barrett L. 2000. *The Experience-Sampling Program (ESP)*. <http://www2.bc.edu/~barretli/esp/>
- Bentler PM 1995. *EQS Structural Equations Program Manual*. Encino, CA: Multivariate Software
- Bolger N, DeLongis A, Kessler R, Schilling E. 1989. Effects of daily stress on negative mood. *J. Personal. Soc. Psychol.* 57:808–18
- Bolger N, Eckenrode J. 1991. Social relationships, personality, and anxiety during a major stressful event. *J. Personal. Soc. Psychol.* 61:440–49
- Bolger N, Zuckerman A. 1995. A framework for studying personality in the stress process. *J. Personal. Soc. Psychol.* 69:890–902
- Bolger N, Zuckerman A, Kessler RC. 2000. Invisible support and adjustment to stress. *J. Personal. Soc. Psychol.* 79:953–61
- Borsboom D, Mellenbergh GJ, van Heerden J. 2002. The theoretical status of latent variables. *Psychol. Rev.* In press
- Bower GH. 1981. Mood and memory. *Am. Psychol.* 36:129–48
- Bryk AS, Raudenbush SW, Congdon RT. 1996. *HLM: Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*. Chicago: Scientific Software Int.
- Clark LA, Watson D, Leeka J. 1989. Diurnal variation in the positive affects. *Motiv. Emot.* 13:205–34
- Csikszentmihalyi M, Larson R. 1992. Validity and reliability of the Experience Sampling Method. See deVries 1992, pp. 43–57
- Dabbs JM Jr, Strong R, Milun R. 1997. Exploring the mind of testosterone: a beeper study. *J. Res. Personal.* 31:577–87
- Dermer M, Berscheid E. 1972. Self-report of arousal as an indicant of activation level. *Behav. Sci.* 17:420–29
- deVries MW, ed. 1992. *The Experience of Psychopathology: Investigating Mental Disorders in Their Natural Settings*. New York: Cambridge Univ. Press
- Diggle P, Liang KY. 2001. *Analyses of Longitudinal Data*. New York: Oxford Univ. Press

- Dohrenwend BS, Krasnoff L, Askenasy AR, Dohrenwend BP. 1978. Exemplification of a method for scaling life events: the PERI Life Events Scale. *J. Health Soc. Behav.* 19:205–29
- Downey G, Purdie V, Schaffer-Neitz R. 1999. Anger transmission from mother to child: a comparison of mothers in chronic pain and well mothers. *J. Marriage Fam.* 61:62–73
- Fabes RA, Eisenberg N. 1997. Regulatory control and adults' stress-related responses to daily life events. *J. Personal. Soc. Psychol.* 73:1107–17
- Fahrenberg J, Myrtek M, eds. 1996. *Ambulatory Assessment: Computer Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies*. Seattle, WA: Hogrefe & Huber
- Feldman LA. 1995. Valence focus and arousal focus: individual differences in the structure of affective experience. *J. Personal. Soc. Psychol.* 69:153–66
- Feldman Barrett L, Barrett DJ. 2001. An introduction to computerized experience sampling in psychology. *Soc. Sci. Comput. Rev.* 19:175–85
- Feldman Barrett L, Jamner LD. 2000. *Capturing the vicissitudes of life: electronic experience sampling*. Workshop presented at Annu. Meet. Soc. Personal. Soc. Psychol., 1st, Nashville, TN
- Gleason MEJ, Bolger N, Shrout P. 2001. *The effects of research design on reports of mood: comparing daily diary, panel, and cross-sectional designs*. Presented at Annu. Meet. Soc. Personal. Soc. Psychol., 2nd, San Antonio, TX
- Gollob HF, Reichardt CS. 1987. Taking account of time lags in causal models. *Child Dev.* 58:80–92
- Gonzalez R, Griffin D. 1999. The correlation analysis of dyad-level data in the distinguishable case. *Pers. Relat.* 6:449–69
- Greenier KD, Kernis MH, McNamara CW, Waschull SB, Berry AJ, et al. 1999. Individual differences in reactivity to daily events: examining the roles of stability and level of self-esteem. *J. Personal.* 67:185–208
- Hackel LS, Ruble DN. 1992. Changes in the marital relationship after the first baby is born: predicting the impact of expectancy disconfirmation. *J. Personal. Soc. Psychol.* 62:944–57
- Hox J. 2002. *Multilevel Analyses: Techniques and Applications*. Mahwah, NJ: Erlbaum
- Jamison RN, Raymond SA, Levine JG, Slawsby EA, Nedeljkovic SS, et al. 2001. Electronic diaries for monitoring chronic pain: 1-year validation study. *Pain* 91:277–85
- Jamner LD, Shapiro D, Alberts J. 1998. Mood, blood pressure, and heart rate: strategies for developing a more effective ambulatory mood diary. In *Technology and Methods in Behavioral Medicine*, ed. DS Krantz, A Baum, pp. 195–220. Mahwah, NJ: Erlbaum
- Jensen-Campbell LA, Graziano WG. 2000. Beyond the school yard: relationships as moderators of daily interpersonal conflict. *Personal. Soc. Psychol. Bull.* 26:923–35
- Jöreskog KG, Sörbom D. 1996. *LISREL 8 User's Reference Guide*. Chicago: Scientific Software Int.
- Kashy DA, Kenny DA. 2000. The analysis of data from dyads and groups. In *Handbook of Research Methods in Social Psychology*, ed. HT Reis, CM Judd, pp. 451–77. New York: Cambridge Univ. Press
- Kennedy JK, Bolger N, Shrout PE. 2002. Witnessing interparental psychological aggression in childhood: implications for daily conflict in adult intimate relationships. *J. Personal.* In press
- Kenny D, Kashy D, Bolger N. 1998. Data analysis in social psychology. In *Handbook of Social Psychology*, ed. D Gilbert, S Fiske, G Lindzey, pp. 233–65. New York: McGraw-Hill
- Kenny DA, la Voie L. 1985. Separating individual and group effects. *J. Personal. Soc. Psychol.* 48:339–48
- Kenny DA, Zautra A. 2001. Trait-state models for longitudinal data. In *New Methods for the Analysis of Change*, ed. A Sayer, LM Collins, pp. 243–63. Washington, DC: Am. Psychol. Assoc.
- Kenny DA, Zautra A. 1995. The trait-state-error

- model for multiwave data. *J. Consult. Clin. Psychol.* 63:52–59
- Larsen RJ. 1990. Spectral analysis of longitudinal data. In *Statistical Methods in Longitudinal Research: Time Series and Categorical Longitudinal Data. Vol. 2: Statistical Modeling and Decision Science*, ed. A von Eye, pp. 319–50. San Diego, CA: Academic
- Larsen RJ, Kasimatis M. 1990. Individual differences in entrainment of mood to the weekly calendar. *J. Personal. Soc. Psychol.* 58:164–71
- Larson R, Csikszentmihalyi M. 1978. Experiential correlates of time alone in adolescence. *J. Personal.* 46:677–93
- Larson R, Richards M. 1994. *Divergent Realities: The Emotional Lives of Mothers, Fathers, and Adolescents*. New York: Basic Books
- Laurenceau JP, Feldman Barrett L, Pietromonaco PR. 1998. Intimacy as an interpersonal process: the importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. *J. Personal. Soc. Psychol.* 74:1238–51
- Laurenceau JP, Feldman Barrett LA, Rovine MJ. 2002. Intimacy in marriage: a daily-diary and multilevel modeling approach. Submitted
- Litt MD, Cooney NL, Morse P. 1998. Ecological momentary assessment (EMA) with treated alcoholics: methodological problems and potential solutions. *Health Psychol.* 17: 48–52
- McArdle JJ, Epstein D. 1987. Latent growth curves within developmental structural equation models. *Child Dev.* 58:110–33
- McCulloch CE, Searle SR. 2001. *Generalized, Linear, and Mixed Models*. New York: Wiley
- Mendoza-Denton R, Downey G, Purdie V, Davis A, Pietrzak J. 2002. Sensitivity to status-based rejection: implications for African-American college students' college experience. *J. Personal. Soc. Psychol.* In press
- Mohr CD, Armeli S, Tennen H, Carney MA, Affleck G, et al. 2001. Daily interpersonal experiences, context, and alcohol consumption: crying in your beer and toasting good times. *J. Personal. Soc. Psychol.* 80:489–500
- Molenaar PCM. 1985. A dynamic factor model for the analysis of multivariate time series. *Psychometrika* 50:181–202
- Molenaar PCM. 1999. Longitudinal analysis. In *Research Methodology in Social, Life, and Behavioural Sciences*, ed. HJ Ader, GJ Mellenbergh, pp. 143–67. Thousand Oaks, CA: Sage
- Morrison DM, Leigh BC, Gillmore MR. 1999. Daily data collection: a comparison of three methods. *J. Sex Res.* 36:76–81
- Moskowitz DS, Hershberger SL, eds. 2002. *Modeling Intraindividual Variability with Repeated Measures Data: Methods and Applications*. Mahwah, NJ: Erlbaum
- Muthén B, Muthén L. 2002. *Mplus Version 2 User's Guide*. Los Angeles, CA: Muthén & Muthén
- Nesselroade JR, McArdle JJ, Aggen SH, Meyers JM. 2002. Dynamic factor analysis models for representing process in multivariate time-series. See Moskowitz & Hershberger 2002, pp. 235–65
- Nisbett RE, Wilson TD. 1977. Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* 84:231–59
- Paty JA, Kassel JD, Shiffman S. 1992. The importance of assessing base rates for clinical studies: an example of stimulus control of smoking. See deVries 1992, pp. 347–52
- Pennebaker JW, Graybeal A. 2001. Patterns of natural language use: disclosure, personality, and social integration. *Curr. Dir. Psychol. Sci.* 10:90–93
- Perrez M, Reicherts M. 1996. A computer-assisted self-monitoring procedure for assessing stress-related behavior under real life conditions. See Fahrenberg & Myrtek 1996, pp. 51–67
- Perrez M, Schöbi D, Wilhelm P. 2000. How to assess social regulation of stress and emotions in daily family life? A computer-assisted family self-monitoring system

- (FASEM-C). *Clin. Psychol. Psychother.* 7: 326–39
- Perrez M, Wilhelm P, Schoebi D, Horner M. 2001. Simultaneous computer assisted assessment of causal attribution and social coping in families. See Fahrenberg & Myrtek 2001, pp. 25–43
- Prosser R, Rasbash J, Goldstein H. 1996. *MLn User's Guide*. London: Inst. Educ.
- Radloff LS. 1977. The CES-D Scale: A self-report depression scale for research in the general population. *Appl. Psychol. Meas.* 1: 385–401
- Rafaeli E, Revelle W. 1999. *Personality, Motivation, and Cognition (PMC) Diary Program*. <http://personality-project.org/pmc/diary.html>
- Rafaeli E, Revelle W. 2002. Affective synchrony: individual differences in mixed emotions. Submitted
- Raffety BD. 2001. *Interactive Software Interview System (ISIS)*. Email raffety@ccplex.com
- Raudenbush SW, Brennan RT, Barnett RC. 1995. A multivariate hierarchical model for studying psychological change within married couples. *J. Fam. Psychol.* 9:161–74
- Raudenbush SW, Bryk AS. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage. 2nd ed.
- Reardon A, Young MA, Blodgett C. 2001. *Patterns of symptom remission in seasonal affective disorder*. Poster presented at Annu. Meet. Soc. Res. Psychopathol., Madison, WI
- Redelmeier DA, Kahneman D. 1996. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain* 66: 3–8
- Reis HT. 1994. Domains of experience: investigating relationship processes from three perspectives. In *Theoretical Frameworks in Personal Relationships*, ed. R Erber, R Gilmore, pp. 87–110. Mahwah, NJ: Erlbaum
- Reis HT, Gable SL. 2000. Event-sampling and other methods for studying everyday experience. In *Handbook of Research Methods in Social and Personality Psychology*, ed. TH Reis, MC Judd, pp. 190–222. New York: Cambridge Univ. Press
- Reis HT, Shaver P. 1988. Intimacy as an interpersonal process. In *Handbook of Relationships*, ed. SW Duck, pp. 367–389. Chichester, UK: Wiley
- Repetti RL, Wood J. 1997. Effects of daily stress at work on mothers' interactions with preschoolers. *J. Fam. Psychol.* 11:90–108
- Rhodewalt F, Madrian JC, Cheney S. 1998. Narcissism, self-knowledge organization, and emotional reactivity: the effect of daily experiences on self-esteem and affect. *Personal. Soc. Psychol. Bull.* 24:75–87
- Roberts NA, Levenson RW. 2001. The remains of the workday: impact of job stress and exhaustion on marital interaction in police couples. *J. Marriage Fam.* 63:1052–67
- Ross M. 1989. Relation of implicit theories to the construction of personal histories. *Psychol. Rev.* 96:341–57
- SAS Inst. 1999. *SAS/STAT User's Guide, Version 7*. Cary, NC: SAS Inst.
- Sayer AG, Collins LM, eds. 2001. *New Methods for the Analysis of Change*. Washington, DC: Am. Psychol. Assoc.
- Searles JS, Helzer JE, Walter DE. 2000. Comparison of drinking patterns measured by daily reports and timeline follow back. *Psychol. Addict. Behav.* 14:277–86
- Shapiro D, Goldstein IB. 1998. Wrist actigraph measures of physical activity level and ambulatory blood pressure in healthy elderly persons. *Psychophysiology* 35:305–12
- Shiffman S. 2000. Real-time self-report of momentary states in the natural environment: computerized ecological momentary assessment. In *The Science of Self-reports: Implications for Research and Practice*, ed. AA Stone, JS Turkkan, CA Bachrach, JB Jobe, HS Kurtzman, VS Cain, pp. 277–96. Mahwah, NJ: Erlbaum
- Shiffman S, Hufford M, Hickcox M, Paty JA, Gnys M, et al. 1997. Remember that? A comparison of real-time versus retrospective

- recall of smoking lapses. *J. Consult. Clin. Psychol.* 65:292–300
- Shiffman S, Stone AA. 1998. Ecological momentary assessment: a new tool for behavioral medicine research. In *Technology and Methods in Behavioral Medicine*, ed. DS Krantz, A Baum, pp. 117–31. Mahwah, NJ: Erlbaum
- Singer JD. 1998. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *J. Educ. Behav. Stat.* 23:323–55
- Singer JD. 2002. Fitting individual growth models using SAS PROC MIXED. See Moskowitz & Hershberger 2002, pp. 135–70
- Snijders TAB, Bosker RJ. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage
- Stonborough THW. 1942. The continuous consumer panel: a new sampling device in consumer research. *Appl. Anthropol.* 2:37–41
- Stone AA, Broderick JE, Kaell AT, DelesPaul PAEG, Porter LE. 2000. Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritis? *J. Pain* 1:212–17
- Stone AA, Schwartz JE, Neale JM, Shiffman S, Marco CA, et al. 1998. A comparison of coping assessed by ecological momentary assessment and retrospective recall. *J. Personal. Soc. Psychol.* 74:1670–80
- Stone AA, Shiffman S. 1994. Ecological momentary assessment (EMA) in behavioral medicine. *Ann. Behav. Med.* 16:199–202
- Stone AA, Shiffman S. 2002. Capturing momentary, self-report data: a proposal for reporting guidelines. *Ann. Behav. Med.* 24:236–43
- Suedfeld P, Pennebaker JW. 1997. Health outcomes and cognitive aspects of recalled negative life events. *Psychosom. Med.* 59:172–77
- Suls J, Wan CK, Blanchard EB. 1994. A multilevel data-analytic approach to evaluation of relationships between daily life stressors and symptomatology: patients with irritable bowel syndrome. *Health Psychol.* 13:103–13
- Swim JK, Hyers LL, Cohen LL, Ferguson MJ. 2001. Everyday sexism: evidence for its incidence, nature, and psychological impact from three daily diary studies. *J. Soc. Issues* 57:31–53
- Tennen H, Suls J, Affleck G. 1991. Personality and daily experience: the promise and the challenge. *J. Personal.* 59:313–38
- Thomas DL, Diener E. 1990. Memory accuracy in the recall of emotions. *J. Personal. Soc. Psychol.* 59:291–97
- Thompson A, Bolger N. 1999. Emotional transmission in couples under stress. *J. Marriage Fam.* 61:38–48
- Tidwell MCO, Reis HT, Shaver PR. 1996. Attachment, attractiveness, and social interaction: a diary study. *J. Personal. Soc. Psychol.* 71:729–45
- Tourangeau R, Rips LJ, Rasinski K. 2000. *The Psychology of Survey Response*. New York: Cambridge Univ. Press
- van Goozen SHM, Wiegant VM, Endert E, Helmond FA, et al. 1997. Psychoendocrinological assessment of the menstrual cycle: the relationship between hormones, sexuality, and mood. *Arch. Sex. Behav.* 26:359–82
- Verbeke G, Molenberghs G. 2000. *Linear Mixed Models for Longitudinal Data*. New York: Springer
- Vinokur AD, Price RH, Caplan RD. 1996. Hard times and hurtful partners: how financial strain affects depression and relationship satisfaction of unemployed persons and their spouses. *J. Personal. Soc. Psychol.* 71:166–79
- West SG, Hepworth JT. 1991. Statistical issues in the study of temporal data: daily experiences. *J. Personal.* 59:609–62
- Wheeler L, Nezlek J. 1977. Sex differences in social participation. *J. Personal. Soc. Psychol.* 35:742–54
- Wheeler L, Reis HT. 1991. Self-recording of everyday life events: origins, types, and uses. *J. Personal.* 59:339–54
- Wilhelm P. 2001. A multi-level approach to

analyze ambulatory assessment data: an examination of family members' emotional states in daily life. In *Progress in Ambulatory Assessment: Computer-assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies*, ed. J Fahren-

berg, M Myrtek, pp. 173–89. Kirkland, WA: Hogrefe & Huber

Willett JB, Sayer AG. 1994. Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychol. Bull.* 116:363–81