CHAPTER 4

Time-Scale-Dependent Longitudinal Designs

THEODORE A. WALLS
WILLIAM D. BARTA
ROBERT S. STAWSKI
CHARLES E. COLLYER
SCOTT M. HOFER

Over the last several years, researchers have employed an increasing variety of intensive longitudinal research designs. The phenomenon under investigation may be a time-varying outcome, such as the number of instances of a given behavior in a given time frame. One question might be whether concurrent continuous or discontinuous factors underlie the manifestation of a behavior. Intensive measures referring to one time scale (such as days) are obtained over the course of a predetermined time frame (which may be measured in terms of weeks or months). The latter is referred to as an epoch or burst. For various scientific or practical reasons, when data on two or more epochs are collected, the researcher decides to incorporate a gap in time between epochs when data are not collected. This approach has been referred to as a "measurement burst design" (e.g., Martin & Hofer, 2004; Nesselroade & Schmidt-McCollam, 2000; Sliwinski, 2008). Examples of the design are appearing in the literature with increasing frequency (e.g., Sliwinski, Almeida, Smyth, & Stawski, 2009; Weinstein, Mermelstein, & Shiffman, 2008).

The term "measurement burst design" appears to have originated in a passage in Nesselroade's (1991) theoretical paper addressing key issues in the study of intraindividual variability. As Nesselroade framed the issue, "longitudinal research designs need to be planned around successive 'bursts' of measurement rather than merely successive ones" (p. 235). As is discussed, the measurement burst design offers certain important advantages to the researcher beyond intensive longitudinal designs that do not vary time scales (c.f., Collins, 2006; Walls & Schafer, 2006). One benefit is the ability to examine a lengthy period of time intensively without obtaining intensive measures continuously over the entire course of the study. A second benefit is the ability to integrate information that is specific to the time scale of measurement. That is, even if the object of one's investigation remains the same, the time scale of measurement will have an important impact on the information that is obtained. For example, if the number of hours a student spends studying is aggregated over the span of a semester, the researcher may fail to observe variability in the number of hours spent studying during the last few days before final exams begin. In this example, data aggregated by semester and data aggregated by week may each provide complementary insights on the phenomenon under investigation, and reveal important differences between students who distribute their studying hours over the semesters and those who "cram" toward the end of the semester.

The measurement burst design is a special case of the interrupted time series design. An interrupted time series design includes a time-registered independent variable such as a planned manipulation(s) or a naturalistic occurrence(s) that results in a change in the parameters describing the series following the time of registration (Campbell & Stanley, 1966; Cook & Campbell, 1979; see also Schmitz, Klug, & Hertel, Chapter 11, this volume). In the burst design, the assignment of measurement periods is based on theoretical and empirically based decisions about the time scale at which a process or phenomenon is optimally captured, as well as consideration of how particular aspects of the process (e.g., intraindividual variance or asymptote) may change over longer periods. From this standpoint, many possible designs could be considered, not only the relatively straightforward design involving equally spaced measurement bursts (e.g., at 6-month intervals), each comprising multiple assessments (e.g., daily for several days). In this chapter, we consider alternative temporal sampling designs within the interrupted time series family and explore other potential designs based on decisions about sufficient or optimal time scales for capturing particular processes. We refer to this framework as time-scale-dependent longitudinal design (TDLD). A key objective is to provide a basis for researchers to develop studies of process that effectively utilize multiple time scales and intensive measurements. The chapter is organized in five sections: central issues, conceptual principles and statistical assumptions, developmental applications, an illustration, and future directions.

Central Issues

The interrupted time series design and its more recent instantiation in the multiple measurement burst study warrant a brief review. Interest in time series designs in a developmental context emerged in the 1930s (cf. Fiske & Rice, 1955).

Origins of the Design: Interrupted Time Series

The interrupted time series (ITS) design was first described by Campbell and Stanley (1963) and was further elaborated in several subsequent publications, including, notably, Campbell (1969a, 1969b) and Cook and Campbell (1979) (see also Schmitz, Klug, & Hertel, Chapter 11, this volume). The original conception of the design involved a single series, such as that resulting from seismic aftershocks over time in one region. They saw potential extensions to the study of causal sequences in social experimentation. The main statistical approach was limited to inspection of differences between epochs in the mean or slope of a univariate series. This was later extended to formal time series analysis and was frequently employed in the classic ABAB design in studies of operant conditioning (Barlow, Nock, & Hersen, 2009; Box & Jenkins, 1970). ABAB refers to the four-step sequence: baseline, treatment, return to baseline, treatment. The purpose of the return-to-baseline step is to verify that any change in the outcome measure coincident with the treatment can plausibly be attributed to treatment as opposed to some extraneous influence.

A sequence of repeated measures is obtained and, at one or more points in the sequence, there is a gap in measurement. The rationale for obtaining multiple pre- and posttest measures is that the criterion variable is likely to exhibit systematic, context-sensitive intraindividual variation. Therefore, in order to determine that a posttreatment increase in the criterion variable is in fact the result of treatment, the researcher will look for an average increase or change in relation to the local maximum. When parameters reflecting information within one period of measurement are considered from period to period, whether the parameters are simple descriptive statistics or more complex indicators of a process, the resulting analysis necessarily occurs at a higher time scale. Hence, a formal definition of a measurement burst design from an ITS perspective is: an ITS in which interest in the evolution of lower time scale processes (such as minute to minute) is considered over higher time scales (such as day to day). By contrast, an approach that simply views bins of minute-to-minute information as equivalent slices without a decision about a higher time-scale-design motivation would fall outside of our definition, or minimally within it. Relevant consideration of the role of time scale in longitudinal studies can be found in Bertenthal (2007) and Newell (1990).

Although it may seem obvious that the measurement burst design is a special case of the ITS, details of the design as a special case have not received careful consideration. Designs in psychophysics (e.g., Swets, 1964) and motor control (e.g., Rosenbaum & Collyer, 1998) may in fact have been early instances of the design, without being explicitly named as such. The motivation for many studies of perceptual and motor processes has been to intensively study the abilities of individuals, and practical considerations naturally lead to multiple sessions or bursts of data collection. Contemporary cases of the design in use have been in extensions of daily diary studies, with epochs of days spaced by several months. Substantively, these studies have focused on "daily hassles" and, to a lesser extent, addiction. The motivation for these studies has often been an interest in micro-level processes in cases in which continuous measurement may have been difficult or impossible. Challenges in these cases include difficulties with participant compliance with burdensome protocols, as well as potential problems with reactivity and validity (Hufford, Shields, Shiffman, Paty, & Balabanis,, 2002; cf. Campbell & Stanley, 1963; Varkey, Pompili, & Walls, 2011).

Campbell and Stanley (1963) stated that "the essence of the time-series design is the presence of a periodic measurement process on some group or individual and the introduction of an experimental change into this time series of measurements, the results of which are indicated by a discontinuity in the measurements recorded in the time series" (p. 37). However, the source of the discontinuity is not always an experimental manipulation. The researcher may, for example, wish to undertake an observational study of naturally occurring events, such as the effect of a flood on local employment (McDowall, McCleary, Meidinger, & Hay, 1980). Such a situation could also be repeatedly interrupted at a calendar time scale, as in the case of sequential events of neighborhood violence, resulting in a naturalistic but nonetheless time-scale-dependent design.

Interest in ITS designs within psychology can be traced back at least as far as the 1930s (cf. Fiske & Rice, 1955). Social and behavioral sciences at this point in time were influenced by learning theory. Learning theorists initially studied organisms' typical responses to overt stimuli before and after exposures to positive and negative reinforcements. As their field progressed, they considered the temporal processes by which complex patterns of behaviors are shaped. Members of the organismic school of developmental psychology were also interested in intraindividual processes. The term *microgenesis*, a translation of the German *Aktualgenese* or "genetic realization," is a concept that can be traced back to Werner's work at the University of Hamburg in the 1920s (Rosenthal, 2004). Microgenesis refers to "the temporal period between the presentation of a stimulus and the formation of a single, relatively stabilized cognitive response to this stimulus" (Flavell & Draguns, 1957, p. 197). This temporal period may be fleetingly small. What distinguishes a microgenetic process is its explicit attention to prior stimulus exposure on subsequent motivational, behavioral, and/or cognitive processes.

In summary, the utility of ITS designs as a means by which to interrupt an ongoing process and thereby draw inferences about that process is not new. However, the recognition that processes may evolve over higher order time scales is a relatively recent innovation from which to devise (or observe) interruption of a process. A few concomitant themes warrant attention in regard to consideration of processes in general; we cover these in the later section on developmental applications.

Conceptual Principles and Statistical Assumptions

Conceptually, the burst designs demand a careful consideration of the processes that are of interest to the researcher. A useful image may be that of a long wall with windows spaced along it and a process going on in a corridor on the other side. The process can be observed (in detail) at the windows, but not between the windows. That is, even if there are gaps in observation, an overall pattern of increase or decline may still be discerned. This is the premise of any longitudinal design, but here emphasis is placed on intensive designs that are capable of achieving a high level of granularity. The benefits of increased granularity include, first, the potential to reveal patterned, recurring temporal shapes embedded within a larger trajectory and, second, an increased ability to detect the presence of discrete, disruptive events that alter the course of the trajectory.

The focus of this chapter includes mapping intraindividual variation and change at two (or more) time scales and the possibility of discontinuous change (i.e., change processes triggered by disruptive events, resulting in a state of disequilibrium) or of stability of processes over longer time intervals. Note that this design is not limited to scenarios in which "days" or within-day processes in the lives of individuals are the smallest unit of analysis. One could look instead at weekly or yearly bursts and at the behavior of cohorts or other aggregated groupings. The sampling of different temporal units will yield different patterns of within-person variation and change that result from different internal states and influences on the individual (e.g., Bertenthal, 2007; Boker, Molenaar, & Nesselroade, 2009; Martin & Hofer, 2004).

Naturalistic Time-Scale-Dependent Time Series

This design reflects processes at different time scales or the evolution of a process studied at one time scale when measured at time points over a higher time scale.

[O1 O2 O3 ...] DELAY [O1 O2 O3] DELAY [O1 O2 O3] . . . .

Note: O = Observations; Observations may be evenly spaced (fixed interval), or they may be unevenly spaced, as in event or location contingent measurements.

Experimental Time-Scale-Dependent Time Series

An elaboration of this design involves study of change in processes via an experiment or randomization to groups, where X denotes the intervention point on the longer time scale:

R G1 [O1 O2 O3 ...] DELAY [O1 O2 O3] X DELAY [O1 O2 O3] . . .
G2 [O1 O2 O3 ...] DELAY [O1 O2 O3] DELAY [O1 O2 O3] . . . .

Another elaboration of the design can be interruption within the lower time scale, such as when traditional ITS designs are replicated over time:

[O1 X O2 O3 . . .] DELAY [O1 X O2 O3] DELAY [01 X O2 O3] . . . .

Whether the design is naturalistic or experimental, the spacing of measurements can be even or uneven. Uneven spacing can be produced either by event-driven measurements or the experimenter's randomization. Finally, of course, additional levels of time, additional groups, or interruptions/interventions at multiple time scales can be employed. This prospect may become increasingly plausible as lower time-scale information on biological and motoric signals is tracked continuously, such as when blocks or nested blocks of time for contexts, marker variables, and developmental epochs are considered.

An interesting question is whether Cook and Campbell's (1979) classic enumeration of threats to internal and external validity for experimental designs is complete with respect to this elaborated form of the design or whether other issues arise. For example, in comparing multiple epochs in which intensive measures have been obtained, a risk is that the passage of time will allow the participant to change to the extent that he or she will answer the same questions through a "different lens," or calibrate his or her responses differently. Whereas practical benefits of the design (e.g., reduced participant burden, possibly lower reactivity, reduced costs) are compelling, further questions can be raised in the following regards. First, in the case of a naturalistic multiple time point interruption, is it clear that the interruptions are occurring purely as a result of the criterion selection variable, or could other factors confound this variable? For example, neighborhood violence could be highly correlated with political or economic shifts, and these changes may remain unchanged even after violent behavior has subsided. Second, can the shape of the primary outcome be assessed well over short time bursts, and is this shape likely to be comparable from epoch to epoch as the overall higher order time course influences phenomena? For example, the epochs reflecting the shape of electrocardiogram signals before, during, and after some calendar time level disease onset, treatment, and remission may be very different both theoretically and in practice. Statistical tests of differences from epoch to epoch of differently shaped phenomena are likely to be very challenging to deploy.

Third, some concerns about selectivity in the case of the experimental time-scale-dependent design are worth considering. The nature of time scale as a criterion for selection of observation/planned missingness periods can be a double-edged sword. On one hand, reducing the number of intensive reporting intervals should reduce reporting burdens and thereby increase self-report validity. On the other hand, temporal undersampling may miss important features of all or some intraindividual change and variation or the events that cause such variation, as some individuals may manifest rapid or other interesting changes between epochs that would be undetected. This is a particularly important consideration when inference to particular individuals is of interest.

From a population perspective, it may not be necessary to capture the process fully in each individual, within overall inferences drawing from all available data and assuming exchangeability. Such a design is one of planned missingness with respect to the time element (e.g., McArdle, 1994). These concerns lead to the need for consideration of contemporary longitudinal concepts regarding time series. For example, an expectation that a series will reflect stationarity would enable the process to be modeled with a number of informative multivariate within-subject approaches, such as dynamic factor analysis or state-space modeling (Chow, Ho, Hamaker, & Dolan, 2010; Zhang, 2010). By contrast, if a specific pattern is expected within each burst, nonparametric functional forms or diverse parameterizations may be needed (Fok & Ramsay, 2006). In this case, the comparability from burst to burst can become much more challenging. Our later example in this chapter handles levels of time scale in the case of pain measurements with multilevel modeling; substantive questions levied at this database at this time did not require consideration of these more challenging modeling issues.

Developmental Applications

The distinction between nomothetic and idiographic approaches has a long history (Allport, 1937; Cattell, 1946, 1966; Lamiell, 1998). Nomothetic researchers seek to delineate broadly generalizable phenomena, whereas idiographic researchers seek to delineate empirically verifiable events and processes that contribute to within-person description and explanation. Although nomothetic research goals can sometimes be achieved using cross-sectional or panel designs, it is difficult to conceive of idiographic research goals that can be achieved without considering the effects of the passage of time. Idiographic methods are generally applied to studying the influence of prior experiences and events on subsequent behaviors and sometimes with continuities or discontinuities of behavior as the individual transitions from one setting to another.

In this tradition, Nesselroade and Jones (Jones & Nesselroade, 1990; Nesselroade & Jones, 1991) proposed what they called a *multivariate, replicated, single-subject, repeated-measures design*. One feature of this design is the use of repeated measures to empirically validate a hypothesized intraindividual temporal sequence. A second feature is the replication of the observed temporal sequence using observations drawn from multiple research participants as a means of establishing generalizability.

Nesselroade and Jones also noted the potential utility of incorporating successive "measurement bursts," giving the example of studies of therapeutic change in the context of interactions between clients and mental health professionals. In this example, a series of intensive measurements can be obtained prior to treatment for comparison with a series of intensive measurements obtained following the completion of treatment. This approach thereby incorporates almost all of the elements of the TDLDs that are the focus of this chapter, except multiple manipulations invoking interruptions.

Research may either focus on a planned interruption consisting of some form of treatment or it may focus on naturalistically occurring events. The latter—for example, fluctuations in daily stress—are, in theory, linked to corresponding fluctuations in behavior. Grzywacz and Almeida (2008) found support for a same-day association between stress and risk of binge drinking both within and between persons; they also found that a "pile up" of stress across at least 3 successive days accounted for additional variability in drinking behavior.

Grzywacz and Almeida's (2008) study may be regarded as an ex post facto design in that the distinction between high stress and low stress days is derived by categorizing the data after they have been collected. It differs from traditional ex post facto designs because the focus is on intraindividual change. This design seeks to simulate quasi-experimental designs by establishing, in retrospect, measures of variables both before and after exposure to the critical stimulus (i.e., "treatment").

Context Dependency

Taking Grzywacz and Almeida's (2008) study as an example, "high stress days" and "low stress days" represent two sets of conditions during which patterns of drinking behavior are expected to differ. The within-day stress–alcohol relationship and the between-day stress–alcohol relationship may be conceptualized as operating at two separate time scales. This leads to a broader point about the range of applications of TDLD.

People who are able to accurately report changes in their social or physical context may nonetheless fail to observe the relationship between changes in context and changes in their own behavior. A researcher may nonetheless be able to note these changes by sampling the individual's behavior in a range of contexts. Hence, diary-based event sampling may mitigate the effects of *consistency bias*, defined as the belief that one's past attitudes and behavior are more consistent with current attitudes and behavior than is actually the case (Jaspers, Lubbers, & De Graaf, 2009). For example, people provide different explanations of their motives for getting divorced when asked before and after the divorces (De Graaf & Kalmijn, 2006).

People overestimate the degree to which their future behavior will match their current intentions. Intentions to practice a health-protective behavior (e.g., safer sex) are at best moderately associated with subsequent behavior (Fishbein, Hennessy, Yzer, & Douglas, 2003). Similarly, *preference reversals* refer to discrepancies between goals and actions; an example is the individual who pays monthly dues at a gym that he or she never attends or who plans to avoid high-calorie snacks but nevertheless fails to resist the urge (Berns, Laibson, & Loewenstein, 2007).

Age-related changes in the ability to accurately identify the contextual determinants of one's own past behavior and to accurately predict one's future behavior remain understudied in the field of developmental psychology. These changes are nonetheless key indicators of an individual's functional capacity to exercise self-regulated behavior.

A benefit of an intensive longitudinal approach, then, is the unique perspective that it affords the researcher. Participants' self-report accuracy is likely more reliable with respect to currently experienced contexts as compared with self-reports of behavior in past or future contexts. A researcher, however, can accumulate "in context" data from multiple contexts and attempt to ascertain the degree to which the same experiences—being offered an alcoholic beverage, having an opportunity to engage in unprotected sex, or more generally having the opportunity to behave in an intention-congruent or intention-incongruent manner—will lead to different outcomes depending on the presence of identifiable contextual cues.

Ecological Constraints and Facilitators

Using TDLD also can afford the opportunity to arrive at more precise understanding of the factors influencing a behavior. For example, if the goal is to test the effect of alcohol use reduction treatments on drinking behavior, it is important to recognize that some individuals—for example, children under close parental supervision—may exhibit reduced drinking that is not attributable to the treatment. Other treatment recipients may face unusually high exposure to cues that elicit drinking behavior. An increase in drinking among these individuals does not necessarily cast doubt on the efficacy or effectiveness of the treatment; rather, it may simply impose a parameter (a certain level of exposure to drinking cues) at which reductions in drinking behavior are more or less likely to occur following treatment. Hypotheses implicating context-sensitive treatment-related change in behavior and hypotheses implicating context-sensitive differences in the maintenance or deterioration of treatment-related behavior change over time are amenable to being evaluated using a TDLD approach.

Classifying Contextual Influences

In documenting the occurrence of transitions by the individual from one context to another, the researcher may be concerned with either *marker* or *process* variables. A marker variable is a discrete point in time at which a psychologically meaningful change has occurred. Examples are marriage, giving birth, being transferred to a new school, and being deployed to combat. A process variable has "antecedents, durations, contexts, and outcomes" (Reese & Smyer, 1983, p. 2). So, for example, the transition to a new job may involve an increase in stress that is either abrupt or gradual, depending on the circumstances, and an eventual decline in stress as the individual adapts. These time-varying factors may be characterized either as dynamically stable or as reflective of developmental change. An example of the former is a hypothetical relationship between stress and impulsive behaviors such as binge drinking. As noted earlier, Grzywacz and Almeida (2008) found a significant difference between days on which high perceived stress was reported and days on which relatively low stress was reported. If the propensity to engage in binge drinking episodes is consistent—elevated on high stress days and reduced on low stress days—this illustrates dynamic stability and in a sense demonstrates replicability.

The distinction between a dynamically stable process and a developmental process may depend on the time scale one chooses to adopt. If the window of observation is a week or a month or even a year, the relationship between stress and binge drinking may appear to be stable. However, if the window of observation is several years, one may find that the more frequent binge drinkers exhibit a steady increase in overall alcohol consumption (Maggs & Schulenberg, 2004–2005). In this scenario, collecting intensive data on binge drinking on successive occasions—let us say, intensive data collected on occasions that are separated by 6 months or a year—will provide data on the changes in the day-to-day association between stress and binge drinking that arise over the course of several months. It is plausible that early in the onset of problem drinking behavior, stress is related to drinking, but that as drinking becomes more habitual and routine, the association with stress declines.

Examples of Marker Variables

Havighurst (1948) developed the concept of developmental tasks. These refer to the "demands, constraints, and opportunities provided by the social environment" as they influence the individual's goal of achieving optimal adaptation (p. 9). Havighurst theorized that failure to achieve a developmental task—for example, attaining an age-appropriate level of education—would adversely affect the individual's chances of mastering subsequent tasks. This may be represented as a marker variable—for example, nonattainment of a high-school diploma as a predictor of subsequent vocational outcomes.

This basic idea has been extended to encompass sequential discrete events. This is referred to as *cumulative disadvantage* (cf. Ross & Wu, 1996). Formal education is an often cited correlate of healthy lifestyle (e.g., abstinence from smoking, moderate alcohol use, etc.). Yet, as Ross and Wu point out, that healthy lifestyle is already evident prior to enrollment among individuals who attend college. One may speculate, then, that the protective effects of education will affect individuals differently depending on their prior status. The sociodemographic factors that predict college attendance also predict alcohol use.

If sociodemographic factors are in fact important determinants of behavior, personological accounts of problem drinking, which rely on constructs such as "low self-control," convey a misleading picture. That is, low self-control is not merely an attribute of persons, but is an effect of the individual's exposure to stressful environments. The marker variable, in this case, may consist of the point in time at which a transition from one environment to another has occurred. One may conjecture that improving an individual's environment will lead to reductions in problem behavior—indeed, evidence suggests that altering living environments is a powerful means of reducing the reinstatement of drug use among abstinent individuals (Solinasi, Chauvet, Thiriet, El Rawas, & Jaber, 2008). A key question from the standpoint of treatment planning is to arrive at a better understanding of the likely speed, peak level, and stability of behavior change based on duration of exposure to an improved environment. This is another example of a type of question that can be addressed using a TDLD approach.

Marker Variables in Ecological Momentary Assessment

In the naturalistic study of behavior, researchers sometimes examine the effects of so-called daily hassles on subsequent behavior. Events—for example, having an argument with a spouse or coworker—are used as markers indicating the onset of increased stress or negative affect (NA). Stress and NA are, in turn, recognized to be proximate antecedents of occasions of cigarette smoking, alcohol use, and other health-compromising self-regulatory behaviors. Ecological momentary assessment (EMA) describes a subset of diary studies in which, typically, multiple measures are collected within a single day, and the goal is to assess intraindividual fluctuations in affective and motivational states such as NA and craving. EMA has been facilitated by advances such as handheld computers, which permit the convenient capture of momentary data (Stone & Shiffman, 1994, 2002).

As a practical matter, it is difficult to establish the moment at which a hassle or disturbing event has inflected a change in affective state. The marker event (e.g., an argument with a coworker) may signal the onset of increased NA, or increased NA may have led to the argument. A reciprocal process might exist, such that increasing NA directly contributes to the escalation of the argument and vice versa. The individual providing self-report data may fail to recognize this as it is happening. Hence, situations arise in which markers may be present but are difficult to localize.

A practical solution to this problem is to regard markers not as a single point in time but as a meaningful pattern of successive measurements. Along these lines, Shiffman and Waters (2004) compared different time scales for analyzing EMA data on smoking lapses. They observed that, prior to lapse events, successive measures of within-day NA showed a pattern of increase. So a "marker" may not consist of a single point but instead of a point at which a change in trajectory can be observed across successive observations.

Illustration

In this short example, we wish both to share the form of data that result from a TDLD and to develop a sense of the questions and analyses that can be pursued with the design. Our data, involving burst reports on pain over calendar time, may not be truly developmental in nature; however, the course of pain over the study period is directly analogous to many developmental studies (e.g., vocabulary growth at intervals within a developmental stage). Our model would be deployed in the same manner for typical developmental data.

The data for the current example come from 116 older adults (mean age = 80, SD = 6, range = 66–95, 72% female, mean education = 15 years, SD = 2) who participated in a longitudinal study of health and cognition. Participants completed self-report measures of physical pain and daily stress on each of six daily sessions for a 14-day period. These bursts of six daily measurements were repeated every 6 months for a 2-year period, yielding up to five bursts of 30 daily assessments (see Sliwinski, Smyth, Hofer, & Stawski, 2006, for more details on study specifics).

Individuals were required to rate how much pain they had experienced during the previous 24 hours on a continuous scale ranging from 0 (no pain) to 100 ("the most pain you could imagine"). Additionally, participants completed a version of the Daily Inventory of Stressful Events (DISE; Almeida, Wethington, & Kessler, 2002). The version of the DISE included in this study consisted of a stem, "During the past 24 hours," followed by five questions: (1) Did you have an argument or disagreement with anyone? (2) Did anything else happen that you could have argued or disagreed about, but you decided to let it pass? (3) Did anything happen to a close friend or relative that turned out to be stressful for you? (4) Did anything stressful happen regarding your personal health? (5) Did anything else happen that most people would consider stressful?

Participants rated the severity of each of each of these stressors using a 4-point scale (0 = *not at all* to 3 = *very*). Individuals were then probed regarding the specific content of experienced stressors. For the current example, we used the subjective severity appraisals as our index of daily stress by calculating the average severity of the events reported across each day. Other quantifications of stress could be used in this context (e.g., peak stress) and may make appropriate theoretical sense for certain hypotheses. However, individuals in the current study typically reported only one stressful event per day, rendering peak stress and average daily stress almost perfectly collinear, and considering the benefits of alternative quantifications of stress is beyond the scope of this example.

Applying the Multilevel Model to Time-Scale-Dependent Longitudinal Data

For the purposes of the current illustration, we used multilevel or hierarchical linear modeling (Laird & Ware, 1982; Raudenbush & Bryk, 2002). This analytic framework allowed us to decompose variability in pain ratings, to model shorter- and longer-term trends in the data, and to consider associations between variables at three different levels of analysis (i.e., across days, bursts, and people). An unconditional multilevel model was used to decompose total variability in pain into three components: day-to-day variability, burst-to-burst variability, and between-person variability. A three-level model was fit to the pain data in which level 1 reflected within-person variability across days, level 2 reflected within-person variability across biannual bursts, and level 3 reflected between-persons variability. A pedagogical treatment of this overall approach with a similar three-level model can be reviewed in Walls, Jung, and Schwartz (2006). The first two levels both describe within-person sources of variability, but over different time scales. Variability at level 3 describes between-person variability in level of pain, averaged across all sessions and bursts. The following model served as the basis for these analyses:

$$\text{Pain}_{ijk} = \pi_{0jk} + \pi_{1jk}\text{Session}_{ijk} + e_{ijk} \qquad \text{(Level 1)}$$

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}\text{Burst}_{.jk} + \beta_{02k}\text{Burst*Burst}_{.jk} + U_{0jk} \qquad \text{(Level 2)}$$
$$\pi_{1jk} = \beta_{10k} + U_{1jk}$$

$$\beta_{00k} = \gamma_{000} + \gamma_{001}\text{Age}_{..k} + V_{00k} \qquad \text{(Level 3)}$$
$$\beta_{01k} = \gamma_{010} + V_{01k}$$
$$\beta_{10k} = \gamma_{100} + V_{10k} \qquad \textbf{(4.1)}$$

where Pain_{ijk} is the pain score for session i, burst j, and person k. The variance components for the level 1 residuals, $\text{Var}(e_{ijk})$, and the level 2 residuals, $\text{Var}(U_{0jk})$, both reflect within-person variability, but each with its own cadence. Level 1 residuals reflect within-person day-to-day variability, and level 2 residuals reflect within-person variability observed across 6-month intervals. Variance in the level 3 residuals, $\text{Var}(V_{0k})$, indicates the amount of stable variability (i.e., individual differences) in affect that exists between persons (averaged across all sessions and bursts). Computing the ratio of any single variance to the sum of all three ($\text{Var}(e_{ijk}) + \text{Var}(U_{0jk}) + \text{Var}(V_{00k})$) will provide an index of the ratio of the total variability that exists at each level (i.e., the day level, the burst level, and the person level). Equation 4.1 also allows the session (i.e., testing reactivity) effects to vary randomly from burst to burst (U_{1jk}) and from person to person (V_{10k}). The effect of burst (or long-term change) is allowed to vary randomly from person to person (V_{01k}). There are also fixed, or average, effects that reflect the average pain score (γ_{000}), the average session effect (γ_{100}) and the average burst effect (γ_{010}). The fixed effect of baseline age (γ_{001}) was also estimated. Age was centered at 80, which was the average age at the first assessment. Estimates of parameters were estimated using SAS Proc Mixed under full information maximum likelihood using all available data (under the assumption of missing at random) with all random effects (i.e., variances and covariances) freely estimated.

Additionally, we examined the associations between daily stress and pain reports at the three levels of analysis—within person across days, within person across bursts, and between persons. Three variables were created to include in the statistical models, corresponding to the stress effects at each level of analysis. At level 1, the stress × day_{ijk} variable refers to the self-reported stress severity for day i during burst j for person k. At level 2, the stress × burst_{jk} variable refers to the average daily stress for person k during burst j, and at level 3, the stress × person_k variable represents the average daily stress for person k aggregated across all days and bursts. The stress variables were grand-mean centered (e.g., Hoffman & Stawski, 2009) but were not further centered within persons or within bursts. We chose to use the grand-mean-centered raw stress values in order to maintain a consistent meaning of the daily stress values across individuals and bursts. The stress variables, centered in this fashion, can be interpreted as such: stress × day_{ijk} is the average within-person day-level stress slope, stress × $\text{burst}_{.jk}$ is the *difference* between the within-person day-level and within-person burst-level slopes, and stress × $\text{person}_{..k}$ is the *difference* between the within-person burst level and the (between-person) person-level slopes (Snijders & Bosker, 1999). We used the ESTIMATE command from SAS Proc Mixed to produce burst-level (stress × day_{ijk} + stress × $\text{burst}_{.jk}$) and person-level (stress × day_{ijk} + stress × $\text{burst}_{.jk}$ + stress × $\text{person}_{..k}$) reactivity slope estimates and their standard errors.

Within this statistical framework, we are able to examine associations at three levels of analysis (i.e., within persons across days, within persons across bursts, and between persons) simultaneously. The association within person across days indicates the extent to which pain ratings were higher on days individuals rated experiencing more severe stress. The association within person across bursts indicates the extent to which pain ratings were higher during bursts (6-month epochs) in which individuals rated experiencing more severe stress. The between-person association indicates the extent to which pain ratings were higher for individuals who reported experiencing more severe stress. Thus, the first two types of associations reflect the extent to which stress and pain ratings are coupled within persons over two different temporal intervals (i.e., days and bursts), whereas the last type of association reflects the extent to which individual differences in stress and pain ratings are associated.

Results

The first model we estimated was a fully empty model, simply to decompose variability in daily pain ratings across the three levels of analysis. As can be seen in Table 4.1 (Model 1), the intraclass correlation coefficient (ICC) was .37 (135.44/(135.44 + 92.84 + 135.97)), indicating that 37% of the variability in pain ratings reflected between-persons differences. The remaining 63% of the variability reflected within-person variability across the two different temporal scales, across days (37%) and semiannual bursts (25%). Thus the vast majority of variability in pain ratings reflected within-person variation.

Next, we sought to expand Model 1 to consider cross-sectional and longitudinal aging trends in physical pain reports. This was accomplished by employing equation 4.1. Figure 4.1 displays the mean pain ratings across sessions and bursts for the duration of the study. Visual inspection of Figure 4.1 suggests a couple of noteworthy points. First, there is noticeable variability in pain ratings across days within a burst, with evidence of pain ratings decreasing across the successive 6 days of assessments. Second, there is noticeable variability in pain ratings across bursts (average pain across the six assessments), with some evidence suggesting that pain ratings are increasing across successive bursts of assessment.

TABLE 4.1. Multilevel Model Estimates for Pain across Days, Bursts, and Persons

	Model 1			Model 2			Model 3		
	Estimate	*SE*	*p*	Estimate	*SE*	*p*	Estimate	*SE*	*p*
Fixed effects									
Intercept	14.26	0.98	< .01	9.07	2.14	< .01	5.56	2.17	0.01
Session				–0.82	0.14	< .01	–0.79	0.14	< .01
Burst				4.46	1.46	< .01	4.18	1.45	< .01
Burst*Burst				–0.55	0.24	0.03	–0.47	0.24	0.05
Baseline Age				–0.08	0.03	< .01	–0.02	0.03	0.49
Stress (WP Day)							0.43	0.12	< .01
Stress (WP Burst)							1.57	1.54	0.31
Stress (BP)							2.38	0.45	< .01
Variance components									
Person level									
Intercept	135.44	18.46	< .01	122.48	34.25	< .01	100.36	32.08	< .01
Cov I.S.				–5.69	3.16	0.07	–5.04	3.06	0.09
Session				0.66	0.48	0.09	0.63	0.47	0.09
Cov I.B.				0.11	8.06	0.98	0.76	7.79	0.92
Cov S.B.				0.49	0.77	0.53	0.54	0.76	0.48
Burst				6.27	2.54	< .01	6.64	2.56	< .01
Burst level									
Intercept	92.84	7.69	< .01	105.31	15.97	< .01	103.04	15.89	< .01
Cov I.S.				–7.74	2.89	< .01	–7.74	2.86	< .01
Session				2.04	0.68	< .01	2.05	0.68	< .01
Session level									
Residual	135.97	3.3	< .01	124.55	3.52	< .01	123.95	3.49	< .01

Note. Cov, covariance; I, intercept; S, session; B, burst.

As can be seen in Model 2 in Table 4.1, the fixed effects of the model revealed that pain reports significantly decreased across days within a burst (estimate = –0.82, *SE* = 0.14). Importantly, the burst effect, which captures more durable longitudinal changes in pain ratings, was significant and positive (estimate = 4.46, *SE* = 1.46), indicating that pain ratings were increasing by about 4.5 points per year of aging. Furthermore, the quadratic burst term was also significant and negative (estimate = –0.55, *SE* = 0.24), indicating that the rate of increase in pain ratings was decelerating. In contrast to the evidence for longitudinal increases in pain ratings, the cross-sectional age effect was significant and negative (estimate = –0.08, *SE* = 0.03), indicating that the participants who were older on entry into the study reported having lower pain on average. Thus, the longitudinal and cross-sectional age trends tell opposing stories, with the former indicating that *getting* older is associated with increasing pain, whereas *being* older is associated with decreased pain. We did test whether baseline age moderated the longitudinal burst effect and found no evidence to suggest such

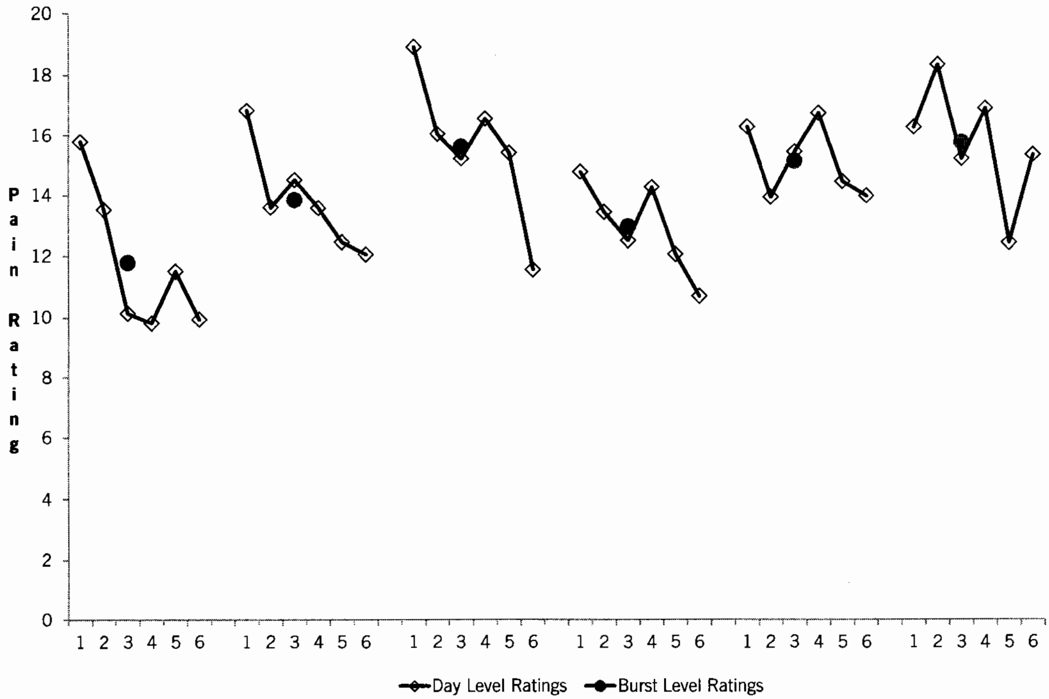

FIGURE 4.1. Average pain levels across days and bursts. Diamonds reflect sample average pain ratings for a given day; circles reflect sample average pain for burst.

moderation. Additionally, it is worth noting that the random effect for burst, which allows for individual differences in the growth trajectory for pain, was significant, indicating evidence of heterogeneity in the rate of change in pain ratings.

The opposing directions of the cross-sectional and longitudinal effects could reflect a selectivity bias, whereby older pain-ridden individuals are underrepresented in the sample, or differential age-related thresholds for pain ratings. Nonetheless, these opposing effects provide strong evidence that the effects of *being* older and of *getting* older are not equivalent. This finding echoes Robinson's (1950) caveat of the ecological fallacy, whereby the use of aggregate statistics of a sample to make inferences about any given individual in the sample can often lead to incorrect conclusions, and indicates that appropriately nuanced and circumspect conclusions regarding cross-sectional versus longitudinal trends and between-person versus within-person associations, given the available data, are necessary.

Finally, we wanted to consider whether daily stress was associated with pain at three levels of analysis (i.e., within-persons across days, within-persons across bursts, and between persons). To test these associations, the stress_day$_{ijk}$, stress_burst$_{.jk}$, and stress_person$_{..k}$ variables were added to equation 4.1 at levels 1, 2, and 3, respectively. The results of these stress analyses can be seen in Model 3 in Table 4.1. At the day level, the stress effect was significant (estimate = 0.43, SE = 0.12), indicating that on days on which participants reported their stressors to be more severe, they also rated their bodily pain on that day as being worse. At the burst level, stressor severity ratings and pain ratings were not significantly associated (estimate = 1.57, SE = 1.54), indicating that there was no reliable association between the average severity of a participant's stressors at a given burst of measurement and his or her average pain levels during that burst. Finally, at the person level, we did see a significant association between stressor severity and pain ratings such that participants who reported their stressors to be more severe on average also reported having worse pain (estimate = 2.38, *SE* = 0.45). Furthermore, we did explore whether the day-level (level 1) or person-level (level 3) stress effects interacted with cross-sectional age or the longitudinal growth trajectories and did not find any such evidence. Thus stressor severity and pain ratings do exhibit significant positive associations between persons, as well as within persons over time. This latter effect, however, appears to be more specific to a faster temporal cadence, day-to-day coupling of stress and pain levels, rather than to a slow temporal cadence, coupling of stress and pain levels over semiannual assessments.

A few important limitations to our illustration should be noted. First, we have not explored the possibility of different error covariance structures carefully for this example. It could be that an autoregressive process of the residuals would be theoretically most appropriate; however, some aspects of the substantive measurements made us cautious about this assumption. Moreover, within the multilevel modeling framework it is not clear how to anticipate a covariance structure for burst-to-burst intervals or possibly higher levels of a TDLD.

Future Directions

This chapter presents a discussion of TDLDs, consisting of specific variations of an ITS design in which time-varying processes are sampled across at least two different time scales. A researcher might wish to examine temporal trends in the amount of time students spend studying during the days or weeks leading up to an exam and compare these trends on successive semesters to determine whether students' studying habits improve over time as a function of increasing commitment, emotional maturity, or the demands imposed by increasingly challenging material. In that the exam constitutes a discrete reference point, it functions as a marker variable, and the amount of time remaining until the day of the exam functions as a process variable that is (one may predict) inversely related to amount of time spent studying. An important feature of these designs is that they permit an evaluation of between-person differences in model parameters obtained within the burst (e.g., time spent studying, pattern of study) that are related to other individual characteristics (e.g., stress, health) and to external outcomes measured within the burst (scores on examination, course grade).

Proceeding from this basic idea, a wide range of practical applications present themselves. One may consider, for example, a habitual response that is dynamically stable at one time scale (e.g., drinking to reduce stress) but that exhibits a directional trend (e.g., increasing alcohol dependence) on another time scale. A related example is the sampling of self-reported quality of marital interactions before and after a treatment aimed at improving communication skills. The value of obtaining preintervention and postintervention data at multiple time points is that instances of marital discord are unevenly distributed over time. Furthermore, retrospective accounts may be biased by the effects of subsequent events on the accuracy of recall.

Used prospectively, a TDLD may offer a useful alternative to a randomized controlled design when randomization is not feasible. If a treatment is structured so that the starting date varies by participant but the duration of treatment is constant, one may observe whether a change that is plausibly attributable to treatment replicates itself for different persons and under varying historical circumstances (Hawkins, Sanson-Fisher, Shakeshaft, D'Este, & Green, 2007). If this multiple baseline approach is merged with TDLD, one may monitor the interaction between treatment dosage and change in treatment-related outcomes, based on comparisons between measures obtained from the same participant before, during, and following treatment. This mitigates many of the threats to validity that have traditionally been addressed using randomization procedures and can rigorously evaluate the question of whether change in behavior has occurred, whether this change is likely attributable to treatment, and whether the change is significant.

Another important feature of the TDLD is that it provides a design-based approach for dealing with reactivity or retest effects, a confound in typical widely spaced longitudinal designs that can limit comparison of results across studies. For example, in the area of cognitive functioning, the application of measurement burst designs permits the modeling of change in asymptotic performance, estimated within each burst, over longer periods of time across burst measurements.

The selection of temporal intervals for a particular application is challenging. The selection of temporal sampling is critical, with oversampling better than undersampling the process of interest (e.g., Boker & Nesselroade, 2002). Replication, generalizability, and comparisons across studies that use different sampling frames will have to be considered, as studies using different sampling intervals are likely to require different interpretations of within- and between-person processes (Martin & Hofer, 2004). Gollob and Reichardt (1987) discuss this issue in terms of measuring causal predictors when they have their maximal effect and the problems in typical studies when this is not a consideration. Also important is the related issue of interval censoring and the temporal resolution by which particular time-varying events can be modeled and compared across studies, although subgroup analyses can be performed on the individuals experiencing a common event. We expect that with increased empirical data, useful temporal sampling frames will be better understood.

The TDLD provides many opportunities to understand individual processes and model the continued development of individuals over time. The design itself is by definition an efficient design in that temporal sampling of different temporal resolutions is performed to capture important aspects of individual functioning across both short-term and long-term periods. There are numerous analytical opportunities for directly modeling within-person processes but also for an increased understanding of between-person differences in a variety of potential parameters that capture the within-person level, change, and variation. The inclusion of more intensive measurements within typical widely spaced longitudinal designs has the potential to enhance measurement of between-person differences but, more important, to permit a basis for understanding long-term change in intraindividual processes.

Acknowledgments

Theodore A. Walls was supported by a research scholar grant from the American Cancer Society (No. MRSG-07-015-01) and Scott M. Hofer was supported by National Institutes of Health Grant No. G026453. We would like to thank Martin Sliwinski for use of the data (supported by National Institutes of Health Grant No. R01 AG12448) that served as the example for this chapter.

References

Allport, G. (1937). *Personality: A psychological interpretation*. New York: Henry Holt.

Almeida, D. M., Wethington, E., & Kessler, R. C. (2002). The Daily Inventory of Stressful Experiences (DISE): An interview-based approach for measuring daily stressors. *Assessment, 9*, 41–55.

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change*. Boston: Pearson.

Berns, G. S., Laibson, D., & Loewenstein, G. (2007). Intertemporal choice: Toward an integrative framework. *Trends in Cognitive Sciences, 11*(11), 482–488.

Bertenthal, B. I. (2007). Dynamical systems: It's all about time. In S. Boker & M. Wenger (Eds.), *Data analytic techniques for dynamical systems* (pp. 1–24). Mahwah, NJ: Erlbaum.

Boker, S. M., Molenaar, P., & Nesselroade, J. R. (2009) Issues in intraindividual variability: Individual differences in equilibria and dynamics over multiple time scales. *Psychology and Aging, 24*, 858–862.

Boker, S. M., & Nesselroade, J. R. (2002). A method for modeling the intrinsic dynamics of intraindividual variability: Recovering the parameters of simulated oscillators in multi-wave panel data. *Multivariate Behavioral Research, 37*, 127–160.

Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Campbell, D. T. (1969a). From description to experimentation: Interpreting trends as quasi-experiments. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 212–242). Madison: University of Wisconsin Press.

Campbell, D. T. (1969b). Reforms as experiments. *American Psychologist, 24*, 409–429.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago: Rand McNally.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasiexperimental designs for research*. Chicago: Rand McNally.

Cattell, R. B. (1946). Personality structure and measurement: I. The operational determination of trait unities. *British Journal of Psychology, 36*, 159–174.

Cattell, R. B. (1966). The data box: Its ordering of total resources in terms of possible relational systems. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (p. 67–128). Chicago: Rand McNally.

Chow, S.-M., Ho, M. R., Hamaker, E. L., & Dolan, C. V. (2010). Equivalence and differences between structural equation modeling and state-space modeling techniques. *Structural Equation Modeling, 17*, 303–332.

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology, 57*, 505–528.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

De Graaf, P. M., & Kalmijn, M. (2006). Divorce motives in a period of rising divorce: Evidence from a Dutch life-history survey. *Journal of Family Issues, 27*, 483–505.

Fishbein, M., Hennessy, M., Yzer, M., & Douglas, J. (2003). Can we explain why some people do and some people do not act on their intentions? *Psychology, Health, and Medicine, 8*, 3–18.

Fiske, D. W., & Rice, L. (1955). Intra-individual response variability. *Psychological Bulletin, 3*, 217–250.

Flavell, J. H., & Draguns, J. (1957). A microgenetic approach to perception and thought. *Psychological Bulletin, 54*(3), 197–217.

Fok, C.C.-T., & Ramsay, J. O. (2006). Periodic trends, non-periodic trends and their interactions in longitudinal or functional data. In T. A. Walls & J. L. Schafer (Eds.), *Models for intensive longitudinal data* (pp. 109–122). New York: Oxford University Press.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58,* 80–92.

Grzywacz, J. G., & Almeida, D. M. (2008). Stress and binge drinking: A daily process examination of stressor pile-up and socioeconomic status in affect regulation. *International Journal of Stress Management, 15*(4), 364–380.

Havighurst, R. J. (1948). *Developmental tasks and education.* New York: McKay.

Hawkins, N. G., Sanson-Fisher, R. W., Shakeshaft, A., D'Este, C., & Green, L. W. (2007). The multiple baseline design for evaluating population-based research. *American Journal of Preventative Medicine, 33*(2), 162–168.

Hoffman, L., & Stawski, R. S. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development, 6,* 97–120.

Hufford, M. R., Shields, A. L., Shiffman, S., Paty, J., & Balabanis, M. (2002). Reactivity to ecological momentary assessment: An example using undergraduate problem drinkers. *Psychology of Addictive Behaviors, 16*(3), 205–211.

Jaspers, E., Lubbers, M., & De Graaf, N. D. (2009). Measuring once twice: An evaluation of recalling attitudes in survey research. *European Sociological Review, 25*(3), 287–301.

Jones, C. J., & Nesselroade, J. R. (1990). Multivariate, replicated, single-subject designs and P-technique factor analysis: A selective review of intraindividual change studies. *Experimental Aging Research, 16,* 171–183.

Laird, N., & Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics, 38,* 963–974.

Lamiell, J. T. (1998). "Nomothetic" and "idiographic": Contrasting Windelband's understanding with contemporary usage. *Theory and Psychology, 8,* 23–38.

Maggs, J. L., & Schulenberg, J. E. (2004–2005). Trajectories of alcohol use during the transition to adulthood. *Alcohol Research and Health, 28*(4), 195–201.

Martin, M., & Hofer, S. M. (2004). Intraindividual variability, change, and aging: Conceptual and analytical issues. *Gerontology, 50,* 7–11.

McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research, 29*(4), 409–454.

McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A. (1980). *Interrupted time series analysis.* Beverly Hills, CA: Sage.

Nesselroade, J. R. (1991). The warp and the woof of the developmental fabric. In R. M. Downs & L. S. Liben (Eds.), *Visions of aesthetics, the environment and development: The legacy of Joachim F. Wohlwill* (pp. 213–240). Hillsdale, NJ: Erlbaum.

Nesselroade, J. R., & Jones, C. J. (1991). Multi-modal selection effects in the study of adult development: A perspective on multivariate, replicated, single-subject, repeated measures. *Experimental Aging Research, 11,* 21–27.

Nesselroade, J. R., & Schmidt-McCollam, K. M. (2000). Putting the process in developmental processes. *International Journal of Behavioral Development, 24,* 295–300.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

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

Reese, H. W., & Smyer, M. A. (1983). The dimensionalization of life events. In E. J. Callahan & K. McCluskey (Ed.), *Life-span developmental psychology: Non-normative life events.* New York: Academic Press.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15,* 351–357.

Rosenbaum, D. A., & Collyer, C. E. (Eds.). (1998). *Timing of behavior: Neural, psychological, and computational perspectives.* Cambridge, MA: MIT Press.

Rosenthal, V. (2004). Microgenesis, immediate experience, and visual processes in reading. In A. Carsetti (Ed.), *Seeing, thinking and knowing: Meaning and self-organisation in visual cognition and thought* (pp. 221–244). Norwell, MA: Kluwer.

Ross, C. E., & Wu, C.-L. (1996). Education, age, and the cumulative advantage in health. *Journal of Health and Social Behavior, 37*(1), 104–120.

Shiffman, S., & Waters, A. J. (2004). Negative affect and smoking lapses: A prospective analysis. *Journal of Consulting and Clinical Psychology, 72*, 192–201.

Sliwinski, M. J. (2008). Measurement-burst designs for social health research. *Social and Personality Psychology Compass, 2*, 245–261.

Sliwinski, M. J., Almeida, D. M., Smyth, J. M., & Stawski, R. S. (2009). Intraindividual change and variability in daily stress processes: Findings from two measurement-burst diary studies. *Psychology and Aging, 24*, 828–840.

Sliwinski, M. J., Smyth, J. M., Hofer, S. M., & Stawski, R. S. (2006). Intraindividual coupling of daily stress and cognition. *Psychology and Aging, 21*(3), 545–557.

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

Solinasi, M., Chauvet, C., Thiriet, N., El Rawas, R., & Jaber, M. (2008). Reversal of cocaine addiction by environmental enrichment. *Proceedings of the National Academy of Sciences of the United States of America, 105*(44), 17145–17150.

Stone, A. A., & Shiffman, S. (Eds.). (1994). Ecological momentary assessment (EMA) in behavioral medicine [Special issue]. *Annals of Behavioral Medicine, 16.*

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24*, 236–243.

Swets, J. A. (Ed.). (1964). *Signal detection and recognition by human observers: Contemporary readings*. New York: Wiley.

Varkey, J. P., Pompili, D., & Walls, T. A. (2011). Human motion recognition using a wireless sensor-based wearable system. *ACM/Springer Journal of Personal and Ubiquitous Computing* [Special issue: Sensor-driven Computing and Applications for Ambient Intelligence], pp. 1–14.

Walls, T. A., Jung, H., & Schwartz, J. (2006). Multilevel models and intensive longitudinal data. In T. A. Walls & J. S. Schafer (Eds.), *Models for intensive longitudinal data* (pp. 3–37). New York: Oxford University Press.

Walls, T. A., & Schafer, J. S. (2006). *Models for intensive longitudinal data.* New York: Oxford University Press.

Weinstein, N. D. (2007). Misleading tests of health behavior theories. *Annals of Behavioral Medicine, 33*(1), 1–10.

Weinstein, S. M., Mermelstein, R., & Shiffman, S. (2008). Mood variability and cigarette smoking escalation among adolescents. *Psychology of Addictive Behaviors, 22*(4), 504–513.

Zhang, G. (2010). Dynamic factor analysis with ordinal variables. In S.-M. Chow, E. Ferrer, & F. Hsieh (Eds.), *Statistical methods for modeling human dynamics: An interdisciplinary dialogue.* New York: Taylor & Francis.