Ecological Momentary Assessment

Saul Shiffman,[1] Arthur A. Stone,[2]
and Michael R. Hufford[3]

[1] Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania
15260; email: Shiffman@pitt.edu

[2] Psychiatry and Behavioral Sciences Department, State University of New York,
Stony Brook, Stony Brook, New York 11794-8790; email: Arthur.stone@sunysb.edu

[3] Cypress Bioscience, Inc., San Diego, California 92121;
email: mhufford@cypressbio.com

Key Words

diary, experience sampling, real-time data capture

Abstract

Assessment in clinical psychology typically relies on global retro-
spective self-reports collected at research or clinic visits, which are
limited by recall bias and are not well suited to address how behav-
ior changes over time and across contexts. Ecological momentary
assessment (EMA) involves repeated sampling of subjects' current
behaviors and experiences in real time, in subjects' natural envi-
ronments. EMA aims to minimize recall bias, maximize ecological
validity, and allow study of microprocesses that influence behavior in
real-world contexts. EMA studies assess particular events in subjects'
lives or assess subjects at periodic intervals, often by random time
sampling, using technologies ranging from written diaries and tele-
phones to electronic diaries and physiological sensors. We discuss
the rationale for EMA, EMA designs, methodological and practical
issues, and comparisons of EMA and recall data. EMA holds unique
promise to advance the science and practice of clinical psychology
by shedding light on the dynamics of behavior in real-world settings.

Contents

INTRODUCTION

Clinical psychologists, along with behavioral, social, and health scientists and practitioners of every stripe, are interested in people's everyday real-world behavior. This interest is perhaps especially marked for clinical psychologists because psychopathology and its functional impairments are expressed in real-world settings: No one is diagnosed or treated because of how they behave in a laboratory or consulting room. Yet, behavior is seldom studied, assessed, or observed as it unfolds in the real world. Instead, both clinicians and researchers rely on global, summary, or retrospective self-reports of behavior: We ask patients how often they experience anxiety, on

average, how many panic attacks they had during the past week or month, how intense their pain generally is during the day, or how depressed their mood has been. Moreover, the emphasis on global assessments can keep us from seeing and studying dynamic changes in behavior over time and across situations, from appreciating how behavior varies, and is governed, by context, and from understanding cascades of behavior, or interactions with others or with our environments that play out as a sequence of events over time. Thus, our frequent reliance on global, retrospective reports seriously limits our ability to accurately characterize, understand, and change behavior in real-world settings and misses the dynamics of life as it is lived, day-to-day, hour by hour. In this review, we discuss an alternative to static retrospective reports—Ecological Momentary Assessment (EMA, Stone & Shiffman 1994), which allows subjects and patients to report repeatedly on their experiences in real-time, in real-world settings, over time and across contexts.

WHAT IS ECOLOGICAL MOMENTARY ASSESSMENT?

EMA is not a single research method; it encompasses a range of methods and methodological traditions, which we discuss below. However, to provide a quick sense of how EMA can be used to inform our understanding of clinical psychology, and as a reference for its paradigmatic characteristics, we describe a prototypical EMA study. Our example is a study of cigarette smoking cessation and relapse (see Shiffman 2005). Smoking is a good target for EMA, as it involves a behavior with clearly discernible small-scale events. Tracking experience over time allows the research to track the process of quitting and relapsing over time. In this study, smokers who had recently quit were asked to monitor their cigarette craving, nicotine withdrawal symptoms, mood, and activities over several weeks, using palm-top computers as electronic diaries. Since episodes of smoking ("lapses")

were of key interest, subjects were asked to record any episodes of smoking as they happened, and were then prompted to complete brief assessments of their craving, mood, and activities during the episode. On top of this, about five times each day, at random times, the electronic diaries also prompted, or "beeped," subjects and administered a similar assessment. These assessments captured not only the events associated with lapses, but the flow of mood, behavior, and events in the hours and days before and after lapses.

These EMA data, captured using the electronic diaries, allowed the investigators to answer a variety of questions (see Shiffman 2005), including: How much craving do smokers experience when they quit smoking? (Surprisingly little, except in discrete episodes; Shiffman et al. 1997a.) Does craving intensity vary with individual characteristics, such as nicotine dependence? (Yes; Shiffman et al. 2004.) Are day-to-day changes in craving intensity associated with variations in subsequent risk of smoking? (Yes, especially craving experienced first thing in the morning; Shiffman et al. 1997a.) Does experiencing emotional distress predispose smokers to lapse? (Yes, but only acute distress, over a period of hours; Shiffman & Waters 2004.) Do situational factors affect lapse risk? (Yes, especially emotional distress, others smoking, and alcohol consumption; Shiffman et al. 1996b.) Do lapses diminish self-efficacy? (Yes; Shiffman et al. 1997b.)

These are just some of the kinds of questions and answers that can be examined using EMA. They illustrate the potential for EMA data to address questions about individual differences, about particular episodes or situations, about the unfolding of processes over time, and about the interactions among these factors. In this way, they illustrate both the richness and the complexity of EMA data.

This study illustrates several key features common to EMA approaches (Stone & Shiffman 1994, Stone et al. 2007a):

■ Data are collected in real-world environments, as subjects go about their

Ecological Momentary Assessment (EMA): methods using repeated collection of real-time data on subjects' behavior and experience in their natural environments

Autobiographical memory: memory processes involved in recalling one's own experience

lives. This is the "ecological" aspect of EMA and allows generalization to the subjects' real lives, i.e., ecological validity.

- Assessments focus on subjects' current state; for example, self-reports ask about current feelings (or very recent ones), rather than asking for recall or summary over long periods. This is the "momentary" aspect of EMA and aims to avoid the error and bias associated with retrospection.
- Moments are strategically selected for assessment, whether based on particular features of interest (e.g., occasions when subjects smoked), by random sampling (to characterize subjects' experiences through representative sampling), or by other sampling schemes.
- Subjects complete multiple assessments over time, providing a picture of how their experiences and behavior varies over time and across situations.

This study is only illustrative. The particular design, assessment schedule, assessment content, and even technology will vary across studies, depending on the behavior researchers are studying, their aims, and their theoretical frameworks. But what EMA studies have in common is the collection of assessments of subjects' current or recent states, sampled repeatedly over time, in their natural environments. With this illustration as background, we now move to highlighting the key aspects of EMA methods.

Momentary, Real-Time Assessment

EMA methods developed in part in response to the limitations of retrospective recall. Although we all feel confident in our own memories, research on autobiographical memory teaches us that memory can be quite unreliable (Bradburn et al. 1987, Tourangeau 2000). Our recollections are not just inaccurate: They are often systematically biased. That is, the errors made in recalling information are not just random noise; rather, they

change the data in systematic ways. For example, people are more likely to retrieve negatively valenced information when they are in a negative mood, thus introducing substantial bias (Clark & Teasdale 1982). Because the dynamics of recall are so important to justifying and structuring EMA methods, these issues are reviewed in more detail below.

Real-World Data

If one is interested in how subjects feel at work, there is no point asking them how they feel in the research clinic—or at home, for that matter. EMA recognizes that many behaviors and experiences can be affected by context. Therefore, in order for assessed experience or behavior to be representative, it has to be sampled in the contexts in which it naturally occurs. Stated more simply, EMA emphasizes ecologically valid observations. Because EMA data are collected in subjects' natural environments—in real life—they should be generalizable to real-world, real-life experience.

Repeated Assessment

EMA studies involve many repeated measures, covering various extents of time with varying intensity of assessment. Some implement a dense schedule of assessment, assessing subjects as often as every 30 minutes (Shapiro et al. 2002) over a period of days. At the other extreme, subjects may be assessed less frequently (e.g., daily) over periods as long as a year (Jamison et al. 2001). Some EMA studies are primarily focused on using these many measures to characterize the subject's "typical" state, aggregating over the repeated assessments to better characterize the subject's average state across situations. More often, EMA studies use the temporal resolution afforded by multiple measures to focus on the within-subject changes in behavior and experience over time and across contexts, addressing how symptoms vary over time or how situational antecedents influence behavior.

In summary, EMA aims to assess the ebb and flow of experience and behavior over time, capturing life as it is lived, moment to moment, hour to hour, day to day, as a way of faithfully characterizing individuals and of capturing the dynamics of experience and behavior over time and across settings.

ECOLOGICAL MOMENTARY ASSESSMENT SCOPE AND HISTORY

EMA is sometimes mistakenly associated narrowly with particular designs, such as the use of randomly scheduled prompts to collect assessments, or with certain technologies, such as the use of palm-top computers. However, EMA is not a single method, much less a particular technology, but rather a collection of methods that share the characteristics described above. EMA includes traditional diaries, whether they use paper and pencil (Green et al. 2006), palm-top computers (Shiffman et al. 1996b), or telephones (Perrine et al. 1995). It encompasses interpersonal interaction diaries (Reis & Wheeler 1991), ambulatory physiological monitoring (Kop et al. 2001), and collection of medication compliance data by instrumented pill bottles (Byerly et al. 2005). The technologies differ, the targets of assessment differ, the schedules of data collection differ, but all of these methods focus on collecting data repeatedly, in close to real time, and in subjects' natural environments. EMA aims to bring these diverse methods under a common framework in order to define higher-order methodological principles, identify commonalities across methods, and thus provide a framework in which researchers and clinicians can select the appropriate methods for their particular research studies or interventions.

Although the term "EMA" was only coined in 1994 (Stone & Shiffman 1994), EMA research, as defined here, has been an active area for decades. A search using the terms "diary," "experience sampling," and "ecological momentary assessment" yielded over 3000 citations in the past 25 years on PsychLit, and 196,000 "hits" on Google™ Scholar. This indicates a substantial volume of research activity using EMA methods.

The emergence of EMA as an important research method is also indicated by a number of books and reviews that have appeared on the topic, including books on EMA methods and findings by Stone et al. (2007b), Hektner et al. (2007), and Fahrenberg & Myrtek (2001); applications to mental health by DeVries (1992); and discussions of data analysis by Walls & Schafer (2006). Other reviews include Wheeler & Reis (1991) on sampling schemes for EMA methods, Scollon et al. (2003) on the promise and challenges of EMA methods, Bolger et al. (2003) on various uses of diary methods, and Piasecki et al. (2007) on applications of EMA methods to clinical treatment. A recent special issue of the *Journal of Personality* (Tennen et al. 2005) described several research programs applying EMA methods to issues relevant to personality, clinical, and health psychology, and several review papers discussed their application to particular domains: Thiele et al. (2002) reviewed the application of EMA methods to clinical psychology, Moskowitz & Young (2006) discussed applications to psychopharmacology, and Beal & Weiss (2003) to industrial psychology. The reader is referred to these sources for more detail than can be included here.

Finally, EMA methods are being used to study a very wide range of behaviors, experiences, and conditions. In a review of published diary studies, Thiele and colleagues (2002) found large groups of studies on pain, mood, anxiety and anxiety disorders, eating, sleep, gastrointestinal disorders, and alcohol consumption. But even this is an incomplete list: EMA studies include studies of depression, social support, initiate relationships, diet, work activity and satisfaction, sexual behavior, psychotherapy, drug use, allergies, psychological stress, adverse effects of medications, self-esteem, and asthma, to name just a few. Clinical disorders studied with EMA

include addictive disorders, eating disorders, anxiety disorders, depression, bipolar disorder, schizophrenia, sexual dysfunction, and ADHD—in other words, the full range of psychopathology. Beyond clinical syndromes and symptoms, EMA is also widely used to study basic adaptation processes relevant to adjustment, such as coping, self-esteem, and social support, as well as behaviors central to health psychology and behavioral medicine, such as coping with illness and treatment, medication compliance, exercise, relaxation, and safe sexual practices, among other behaviors. In sum, EMA methods are used in most domains of concern to clinical psychology.

Historical Roots

EMA draws together several historical traditions, including diaries, self-monitoring, experience sampling, ambulatory monitoring, and others. The oldest is the use of written diaries for research (Verbrugge 1980), which was systematically deployed in clinical research in the 1940s. Self-monitoring of particular behaviors or experiences (Korotitsch & Nelson-Gray 1999) has a long history in research and particularly in behavioral treatment. This includes simple counts of clinically relevant events, but also collection of data about their antecedents and contexts, which provided data for "naturalistic functional analysis" (Schlundt et al. 1985). Related methods included self-monitoring of particular targets; a prominent example is the Rochester Interaction Record (Reis & Wheeler 1991), which subjects used to record every social interaction. Self-monitoring approaches often aimed to capture all relevant events, and as a result did not focus on sampling issues or collect data outside of target events.

Other historical streams that fed into the development of EMA focused on broad descriptions of subjects' behavior. One came from the focus of the Kansas School of Ecological Research (Barker 1978) on continuous observation of behavior through the day in the natural environment. Another derived from the ethnographic method of describing individuals' allocation of time, often to describe differences among societies (Szalai 1966). A variant of this within schools of management and business examined the behavior of workers in the workplace to understand how they used their time and what activities they engaged in.

More central to the development of modern EMA methods was the development by Czikszentmihalyi and colleagues (DeVries 1992, Hektner et al. 2007) of the Experience Sampling Method, demonstrating the innovation of randomly sampling experience, initially using pagers to "beep" people at random times to prompt them to complete diary cards reporting their activity, mood, and/or thoughts. The development of electronic diaries, based upon the emerging technology of handheld computers, opened up new opportunities for more complex and sophisticated EMA protocols that incorporated, combined, and expanded upon the various approaches to collecting EMA self-report data.

Enabled by technological developments, ambulatory monitoring of cardiovascular function, which became possible with the development of wearable cardiac monitors, has been used for several decades as a means of understanding the link between experience and cardiovascular health (Turner et al. 1994). Ambulatory monitoring did not rely on self-report (although it was often accompanied by written diaries) and also enabled continuous or near-continuous recording. Recent developments have expanded physiological monitoring to other parameters, such as galvanic skin response, temperature, motion, and others (Wilhelm et al. 2003). Often little attention was paid to sampling, in part because the monitors could collect data almost continuously (e.g., actigraph or heart-rate recording) or at very high and stable frequencies. Some physiological monitoring devices (e.g., monitoring of blood-glucose, pulmonary function) require subjects to actively make periodic assessments, and thus raise many of the sampling issues that arise with self-report.

Technological developments have also enabled automated EMA assessment of behaviors (e.g., pill taking; Cramer et al. 1989) and even of the physical environment (e.g., air sampling; Saito et al. 2005).

These diverse traditions developed in different disciplinary contexts, often with distinct assessment targets. For example, the Experience Sampling Method focused on subjective states (Hektner et al. 2007 refer to it as systematic phenomenology), whereas self-monitoring focused on behaviors, and ambulatory monitoring focused on physiological parameters. Moreover, because of their particular disciplinary histories and content foci, these methods have been discussed in different literatures, and investigators in one tradition often seem unaware of the others. Yet, these methods share many goals, concerns, and approaches. The object of EMA, which attempts to encompass all of them, is to unify these diverse approaches under a common methodological framework.

AUTOBIOGRAPHICAL MEMORY AND LIMITATIONS OF RECALL

A major motivation for EMA is to avoid the pitfalls and limitations of reliance on autobiographical memory. In this section, we expand on the issues and concerns regarding autobiographical memory, both to explain the rationale for EMA and to provide an essential foundation to considerations of the uses of EMA and specific EMA designs.

Research on autobiographical memory (Bradburn et al. 1987) indicates that recall is not just subject to random error but also is fraught with systematic bias, which can distort recall even after relatively short intervals. Modern cognitive science considers that much of what we "recall" is actually reconstruction, pieced together from fragmentary inputs through the use of various heuristic strategies. Many experiences are not retained in memory, so often the information we are asked to provide simply is not available for direct retrieval. Experiences are particularly

likely to be encoded and retrieved if they are emotionally salient or are unique; routine experience is less likely to be encoded and harder to retrieve. Moreover, the process of retrieval itself is subject to bias because the accessibility of particular content in memory varies with the subject's mental state at the time of retrieval.

Importantly, research inquiries usually ask subjects not to just retrieve but also to aggregate and summarize their experiences (e.g., "How intense was the pain, on average, today?"). When trying to answer such questions, subjects do not recall, enumerate, and then aggregate their experience over time (Bradburn et al. 1987). Rather, they use a variety of heuristics to estimate the answer. The use of cognitive heuristics and the processes of retrieval account for much of the bias in recall data.

A key example of a biasing cognitive heuristic is the "availability heuristic." In its original form, as described by Tversky & Kahneman (1973), this refers to how people make judgments about the frequency of events. When deciding how often an event (say, a fight with a spouse) occurs, a person tries to retrieve an example. If an example is easy to think of (i.e., it is easily "available" in memory), the event is considered to be frequent. One can see how this heuristic would work much of the time—rare events should be harder to "find" in memory. But one can also see how biased it can be: if a fight was particularly memorable because it was intense, if it is easily recalled because it was recent, or because one was recently reminded of it (say, by a previous question in the assessment), then the frequency of fights will be overestimated.

Importantly, the process of memory retrieval is itself subject to bias by the person's context and mental state at the time of recall. It has been shown, for example, that subjects in a negative mood more easily recall negative information than positive (Kihlstrom et al. 2000). Similarly, subjects who are in pain find it easy to remember past pain, but harder to recall pain-free states; accordingly, it's been

shown that subjects who are in pain at the time of recall will overestimate their past pain (Eich et al. 1985). This has profound implications for research because it suggests that the subject's state and situation at time of reporting will influence what is reported.

A dramatic demonstration of the biases in recall—and an indication of how quickly these biases can set in—was reported by Redelmeier et al. (2003). Summary ratings of pain by subjects who had undergone a colonoscopy 20 to 30 minutes earlier were found to be unduly influenced by the peak level of pain (presumably because it was most salient) and the pain intensity at the end of the procedure (most recent). In other words, recall did not accurately represent the average pain over the interval, because it was based on a few of the most memorable moments, essentially ignoring most of the experience. This shows the potential for bias even over short intervals.

Besides being distorted by the operation of heuristic recall strategies, memory is also influenced by what we know and believe rather than actually recall. People unconsciously reorganize their "memories" to make them fit a coherent script or theory of events or to reconcile events with what transpired subsequently (Ross 1989). These biases are particularly pernicious because they tend to produce recalled patterns that are coherent and that may conform to theoretical predictions, even if they are false.

It is important to understand that these biasing processes operate involuntarily and unconsciously. They do not represent distortion by uncooperative or defensive subjects—this is simply the way memory operates. Research inquiries often maximize the potential for bias by asking about routine events, asking subjects to summarize their experience, and soliciting recall in unusual settings (i.e., research laboratories) and contexts (e.g., after being asked other questions) that can bias recall. Thus, heuristics that serve well enough to address the demands placed on autobiographical memory in everyday life can break down when called upon to produce accurate research data.

THE USES OF ECOLOGICAL MOMENTARY ASSESSMENT

EMA data are collected for a variety of purposes (Bolger et al. 2003). We categorize these into four classes: (*a*) characterizing individual differences, (*b*) describing natural history, (*c*) assessing contextual associations, and (*d*) documenting temporal sequences. We illustrate each with an example.

Individual Differences

When used to characterize individual differences, EMA data are aggregated to obtain a measure of the subject that is collapsed across time (i.e., across multiple EMA measures); for example, the average intensity of pain experienced by a pain patient. As an extension of this, aggregated EMA data might be used to quantify subjects' characteristics at two different time points; e.g., pain before and after treatment administration. As estimates of subject characteristics, aggregated EMA data are expected to provide assessments of individuals that are more reliable (because of aggregation) and more valid (because of avoidance of recall bias, representative sampling, and ecological validity). Of course, if the variable is very stable over time, if recall bias were not present, and if contextual factors did not influence the variable, then there would be no advantage in using EMA.

Natural History

To describe natural history, EMA measures are analyzed for trends over time. In this case, the within-subject variation over time itself is the focus, and time is the independent variable, the X-axis in a graphical representation of the data. For example, McCarthy et al. (2006) documented the trajectories of various withdrawal symptoms that smokers experienced after quitting. The EMA data demonstrated that some symptoms peaked immediately when smokers quit and then decreased over time, while others increased and persisted, and still others increased only gradually

over time. These patterns contradicted widely held notions about the course of the withdrawal syndrome and were associated with differences in treatment outcome. Basic descriptive information about the natural history of symptoms over time can often be an important foundation for understanding clinical disorders and outcomes.

Contextual Associations

Studies that examine contextual associations look at the association or interaction between two (or more) phenomena that co-occur in time. Analyses of contextual associations are often cross-sectional, even when data are collected longitudinally, in that they examine the co-occurrence of events or experiences, not their sequence. In these analyses, time is not explicitly represented—it is more of a stage against which the events of interest play out. For example, Myin-Germeys et al. (2001) examined emotions accompanying stressful events as a way to test a diathesis-stress model of schizophrenia. They postulated that vulnerability to schizophrenia would be reflected in excess emotional responses accompanying stress. Schizophrenics, their first-degree relatives (who are genetically vulnerable), and normal controls were assessed 10 times daily about stressful events and mood. An examination of individual differences in average mood showed that the schizophrenics reported more negative affect and more stressful events, whereas vulnerable individuals and normal controls did not differ. But a look at stressor-mood associations revealed that the first-degree relatives reacted more strongly than did controls. Thus, examination of the association between stressors and mood at particular moments was key to understanding what vulnerabilities might be conferred by a genetic predisposition to schizophrenia.

Understanding the momentary cross-sectional associations between different aspects of experience has also been important for foundational studies of the structure of behavior and experience; for example, data

on the covariation of momentary emotions have been central in the debate about whether positive and negative emotions are polar opposites or are independent dimensions and can be experienced simultaneously. Feldman-Barrett & Russell (1998) used EMA data to address the argument that although one could be both happy and distressed over some interval of time, in a particular moment, one could be either happy or distressed but not both.

Although most designs examine associations between different variables within the same person, an interesting variation considers how one person in a relationship affects the other (Bolger & Laurenceau 2005). For example, Larson & Richards (1994) asked members of families to track their experience in parallel and examined how the mood of each affected the other. They found, for example, that a husband's mood when he comes home from work significantly influences his wife's mood, but not vice versa.

Temporal Sequences

Finally, the longitudinal nature of EMA data is used to explicitly examine temporal sequences of events or experiences, to document antecedents or consequences of events or behaviors, or to study cascades of events. In these analyses, unlike those above, the order of events or assessments is explicitly considered and is a key focus. The previously cited study of quitting smoking (Shiffman et al. 1997b) assessed smokers' affect and self-efficacy before and after lapses to smoking, and their effects on subsequent progression toward relapse, to test Marlatt's theory (Curry et al. 1987) that the psychological response to lapses is what drives progression toward relapse. Comparing assessments obtained before the lapse and afterward confirmed the theory's hypothesis that lapses would result in increased negative affect and decreased self-efficacy (Shiffman et al. 1997b). Continued EMA monitoring, however, contradicted the theory's prediction that increases in negative affect and decreases

in self-efficacy would predict the risk of subsequent progression to another lapse or relapse (Shiffman et al. 1996a). Importantly, retrospective analyses had appeared to show a relationship between initial responses to a lapse and progression to relapse (Curry et al. 1987), but later comparisons with EMA data (Shiffman et al. 1997c) showed that retrospective reports of lapse episodes were inaccurate and biased: Subjects recalled their mood as worse than it actually had been, and those who had returned to smoking at the time of recall exaggerated how demoralizing the initial lapse had been. Thus, prospective assessments of the flow of behavior and experience, and of the antecedents and consequences of events, can enable a more valid and more detailed understanding of behavior.

These examples illustrate the use of EMA data to evaluate hypotheses regarding the dynamic interactions among processes over time. Data provided by EMA studies may be likened to a movie, in which dynamic relationships emerge over time, whereas global or recall measures are analogous to a still photograph, a single static snapshot of time. By providing temporal resolution, EMA methods allow investigators to examine sequences of events and experiences and enable them to describe and analyze cascades of events and interactions between events that shape behavior over periods of minutes, hours, or days. Insight into microprocesses—the interplay or cascade of cognitive, affective, and behavioral variables over short intervals of time—is particularly important because many theories of psychopathology and treatment focus on how these processes unfold over time. Evaluation of microprocesses also facilitates development of interventions because an understanding of how affect, cognition, and behavior interact and unfold over time helps identify leverage points for timely, and at least potentially more efficacious, clinical interventions. The ability to shed light on dynamic processes and situational influences is potentially the most critical contribution of EMA methods to clinical psychology.

DATA COMPARING RETROSPECTIVE AND REAL-TIME REPORTS

Research suggests that recall measures, and especially summary measures, are often biased due to the use of mental heuristics to recall information. This implies that there will be discrepancies between EMA-based and recall-based assessments of the same period. Relatively few studies have examined such discrepancies. In discussing this literature, we distinguish comparison with aggregated EMA measures from comparisons of disaggregated time-specific estimates.

Comparison of Aggregated Recall-Based Data and EMA

A number of studies have evaluated the relationship between EMA- and recall-based assessments, comparing EMA assessments averaged over some interval with recall-based measures for the same interval. In some instances, the two methods yield similar estimates (Shrier et al. 2005), and in some other cases, recall methods produce lower estimates of intensity or frequency than do EMA methods (Carney et al. 1998, Litt et al. 2000). However, in many domains, recall-based assessments tend to yield higher estimated levels than diary ratings of the same target events: That is, symptoms tend to be described as more frequent, more intense, and longer lasting, sometimes dramatically so (Broderick et al. 2006, Houtveen & Oei 2007, Shiffman 2007, Shiffman et al. 2006; see review in Van den Brink et al. 2001). Behavior frequency is also often overestimated in recall (Homma et al. 2002, Shiffman & Paty 2003). This phenomenon may be the result of the undue influence of more salient experiences in recall: Intense pain is more salient than no pain, headaches are more salient than non-headaches, and so on, leading recall data to overestimate clinical symptoms. An example of this is the finding that subjects who had the most intense headaches also overestimated

headache frequency (Van den Brink et al. 2001).

Studies have also examined the correlation between aggregated EMA and recall-based data, which assesses whether the ranking of individuals is similar across the two sources, and can be high even when there are differences in the two means. Like the findings on mean differences, the findings from correlational studies are also variable. Studies of drug use report good correspondence for frequency or quantity used (Carney et al. 1998, Shiffman & Paty 2003), but very poor correspondence when characterizing situational patterns of drug use, which involve more complex judgments (Shiffman 1993, Todd et al. 2005). In one study of pain among chronic pain patients (J. Broderick, J. Schwartz, and A.A. Stone, unpublished observations), correlations in the 0.70s were seen between EMA and recall-based measures of pain intensity, whereas Van den Brink (2001) reported correlations around 0.20 for headache frequency, intensity, and duration.

Thus, findings on correspondence between aggregate EMA data and recall-based estimates are inconsistent and variable. One likely reason is that the magnitude and direction of recall bias can differ across subjects and settings. Individual differences can affect recall accuracy. Feldman-Barrett (1997) found that recalled distress was exaggerated among more neurotic subjects and positive affect was exaggerated among extraverted subjects, and Van den Brink (2001) describes other individual differences that moderate recall bias. Beliefs can also moderate recall bias. McFarland et al. (1989) found that women who believed that their menstrual cycle influenced their mood reported exaggerated negative mood in retrospect, but only for menstrual days. In contrast, women who did not believe menstruation influenced mood did not show any bias, either for menstrual or nonmenstrual days.

Variation in the pattern of target symptoms themselves can moderate bias. Stone et al. (2005) showed that subjects whose pain was relatively constant over time were able to estimate their pain more accurately since, in effect, they could easily know their typical pain level or estimate accurately from any given time point. In contrast, subjects whose pain was variable demonstrated considerable bias. Furthermore, Broderick et al. (2006) and others have found that recalled pain was particularly exaggerated when subjects were in pain at the time of recall and understated when they were not.

Thus, the validity of recall data likely varies with characteristics of both the samples and the setting. It is likely that the validity of recall is also influenced by the duration of the recall interval and the variability, salience, and uniqueness of what is being assessed: Some kinds of end points may be well estimated, whereas others are not. Better understanding of these relationships could help establish when EMA is likely to add the most value compared with aggregated recall measures.

Correlations Between Disaggregated EMA and Recall

The studies reviewed above assess the correlation between EMA- and recall-based estimates of experience when EMA data are aggregated over time to broadly characterize person-level effects, and compared with similar data based on global recall. However, EMA data are most often used not just to characterize between-person differences, but also to characterize within-person variations in experience over time, so that the question is whether recall measures are able to accurately reflect time-specific data. The available evidence suggests that they cannot. Even when global recall data correlate well with aggregated EMA data, as in the studies cited above (J. Broderick, J. Schwartz, and A.A. Stone, unpublished observations; Carney et al. 1998; Shiffman & Paty 2003), they do not adequately capture time-varying data, with correlations for recall of specific days hovering in the range of 0.20–0.40. Moreover, analyses show that the cross-day correlation

also varies widely between subjects, with some subjects actually showing negative correlations between their recalled reports and real-time reports of the same target events (Carney et al. 1998, Searles et al. 2000). In other words, even when recall appears adequate to characterize aggregate experience, it is not typically adequate to characterize day-to-day changes in cognitions and behaviors of interest, which are typically the focus of EMA research. This highlights one of the unique contributions of EMA to the study of processes that unfold over time.

Construct Validity of Real-Time Versus Recall Assessments

In addition to comparisons of EMA-based and aggregated recall-based assessments to each other, comparisons can be made based on their construct validity—in other words, their relationships with other theoretically relevant constructs. Several studies illustrate cases in which EMA data show incremental validity over and above retrospective or global assessments of similar constructs. Kamarck and colleagues (2007) directly contrasted the effect of job strain on cardiovascular outcomes when strain was measured by EMA versus global measures. Subjects filled out a standard global job strain questionnaire and used an electronic diary to collect EMA data on experienced strain every 45 minutes for six days (data from the EMA was averaged to create a single EMA summary variable). The study's outcome measure was a prospectively assessed biological outcome: the progression of blockage in the carotid artery (which correlates with blockage of coronary arteries) over the subsequent three years. EMA-based measures of job strain predicted progression of arterial occlusion; however, traditional global questionnaire measures of job strain did not. Moreover, heart rate assessed by EMA in the natural environment also independently predicted progression of carotid blockage, whereas heart rate measured in the lab did not. This illustrates the potential of EMA

data to shed light on relationships that are missed when relying on global retrospective self-reports.

In a study of smoking, Shiffman and colleagues (2007) similarly compared global and EMA-based measures, in this case, measures of "negative-affect smoking"—the tendency to smoke when distressed. The EMA-based measure estimated negative-affect smoking by comparing negative affect on smoking and nonsmoking occasions; the global assessment used standardized and validated questionnaires. Shiffman et al. tested the prediction that smokers engaged in negative-affect smoking would be more vulnerable to relapse after they quit smoking. This was found to be true for EMA-assessed negative-affect smoking, but not for assessments based on standard recall-based questionnaires.

The literature also suggests that EMA measures may sometimes mirror the findings of recall measures, but may capture the target constructs with less noise and greater sensitivity. In a study of the efficacy of analgesics in the treatment of rheumatoid arthritis, Nived and colleagues (1994) reported that diary data differentiated active treatment from control after only four weeks of treatment, whereas it took 24 weeks for the differences to become apparent in recall measures collected at clinic visits. In a related vein, analyses have shown that real-time diary-based methods resulted in decreased error variance when capturing events (McKenzie et al. 2004) and scaled ratings (Pearson 2004), making such data more sensitive to treatment effects.

In contrast to these studies favoring EMA data, J. Broderick, J. Schwartz, and A.A. Stone (unpublished observations) recently found that EMA and recall-based measures of pain were equally correlated with medication-taking and social impairment. Importantly, other studies have illustrated cases in which recall measures were actually better predictors of subsequent behavior than were EMA data. These cases are illuminating. In a study of bias in recall of pain, Kahneman and colleagues (Redelmeier et al. 2003) showed that

participants who were randomized to undergo a longer colonoscopy, with more overall pain, but who experienced less intense pain near the end of the procedure actually recalled lower overall pain. A follow-up revealed that subjects who had undergone the longer procedure—which was experienced as more painful in real time—were more likely to return for a repeat colonoscopy. In other words, their future behavior was related to the retrospective summary of pain, not the momentary experience. Similarly, Oishi & Sullivan (2006) found that break-up of dating couples was better predicted from their retrospective summaries of relationship dissatisfaction than from their daily ratings. These findings make sense, because people shape their subsequent behavior—whether to return for a colonoscopy, whether to proceed with a relationship—by reference to their stored summary memory of the experience, not the actual experience at the time.

These examples illustrate a key point: Even if they do not match momentary experience, retrospective impressions or global beliefs can exercise greater control over subsequent behavior, since they represent the information that people use to make subsequent decisions. In this sense, the momentary and retrospective reports represent different perspectives on the same event. If one is interested in what was experienced in the moment (e.g., to understand the physiological effects of the pain or to evaluate the effects of an analgesic) or wishes to understand what are the proximal determinants of a specific event, then the momentary data are likely to provide a more valid picture. But if one is interested in the person's impressions of an event or in the prediction of future behavior, then the retrospective impressions may well be better predictors than EMA. This suggests that researchers evaluating the validity and utility of EMA and recall data must consider whether they are concerned with understanding the experience as lived or with understanding subjects' impressions of those experiences. As we have argued with regard to assessments of psychological

coping (Stone et al. 1998), confusion about what is being assessed can misdirect theory and research.

The discrepancy between momentary experience and global summary recall also suggests that it would be worthwhile for research to help us understand how momentary experiences are integrated in the development of global judgments. More research is also needed on the circumstances in which global or retrospective judgments are valid and circumstances in which EMA data are necessary. In the interim, EMA is likely to provide unique insights about processes that characterize dynamics of behavior, cognition, and affect over time.

ECOLOGICAL MOMENTARY ASSESSMENT DESIGNS AND APPROACHES

In this section, we describe and categorize EMA designs, by which we mean the scheme that dictates the scheduling, arrangement, and temporal coverage of EMAs. In global assessments such as personality questionnaires, the researcher assumes that the assessment captures the subject's entire experience in one fell swoop, rendering it unimportant when the assessment is made. In EMA, one assesses moments or periods of time, raising the issue of how to ensure that the moments or periods assessed are representative of the subject's experience. In many cases, the assessments can be conceptualized as a sample of the person's experience or behavior. Thus, designing an EMA protocol can essentially amount to designing a sampling scheme for moments in an individual's life. The most important influence on the design must be the aims of the study.

EMA sampling and assessment schemes can be roughly divided into event-based sampling and time-based sampling schemes (Shiffman 2007, Wheeler & Reis 1991). Event-based schemes do not aim to characterize subjects' entire experience, but rather to focus on particular discrete events or

Event-based sampling: a method of data collection whereby a recording is made each time a predefined event occurs

Time-based sampling: a method of data collection whereby a recording is solicited based on a time schedule, often based on random time intervals

episodes in subjects' lives—e.g., headaches (Niere & Jerak 2004) or drinking episodes (Todd et al. 2005)—and organize the data collection around these events. Time-based sampling typically aims to characterize experience more broadly and inclusively—e.g., observing how mood varies over time—without a predefined focus on discrete events.

Event-Based Monitoring

In many disorders, the clinical and research interest is in particular events or episodes, e.g., instances of drinking, violence, or panic attacks. These cases lend themselves to event-based monitoring, in which assessments are triggered by the occurrence of a predefined event of interest to the investigator. For example, subjects might be asked to complete an assessment when they have a panic attack (Taylor et al. 1990), engage in a social interaction lasting more than 10 minutes (Reis & Wheeler 1991), or take a medication (Jonasson et al. 1999). Typically, the subjects themselves determine when the event has occurred and initiate an assessment (though some events can be automatically detected by devices; see Kop et al. 2001). Such protocols require clear definitions of the event, which can be surprisingly thorny to delineate: If subjects are to make a record every time they eat, does chewing gum count as eating? Sometimes, target events are defined as episodic flare-ups of otherwise continuous experiences, for example, a pain episode in which pain is experienced more intensely (McCarberg 2007), or an episode of intense cigarette craving (Shiffman et al. 1997a). Defining the algorithm for declaring an event is particularly difficult—and important—in these cases.

If one only needs a record of events, for example, to ascertain their frequency and time distribution, subjects need only note that an event has occurred, e.g., press a button on a recording device. More often, investigators wish to collect data about the event: its duration, intensity, antecedent mood, etc.

(Schlundt et al. 1985, Shiffman et al. 1996b). If the events are too frequent, it may not be realistic to assess each event, in which case a subset of them can be sampled at random (Shiffman et al. 2002).

An important limitation in the use of event-based monitoring is that there is often no way to independently assess or verify compliance; i.e., there is no way to know whether events occurred that were not entered or (less likely) entries made for events that did not occur. In a few studies, diary entries have been compared to records made by separate electronic devices (e.g., instrumented medication dispensers); these have suggested that medication compliance is exaggerated in self-report (see Hufford 2007). Event entries can sometimes also be roughly confirmed by biochemical measures; e.g., Shiffman & Paty (2003) reported that subjects' electronic diary records of cigarettes were consistent with biochemical measures of smoking. In any case, event reports are subject to error resulting from poor compliance or falsification.

Time-Based Designs

Some clinical phenomena—e.g., mood, pain—vary continuously and are not easily conceptualized in an episodic framework. In some instances (e.g., actigraphy, heart rate, skin conductance), the phenomenon can be monitored continuously. In the more typical case, where this is not possible, EMA protocols rely on time-based sampling. There are many varieties of time-based sampling schemes, which vary in schedule, frequency, and timing (see Delespaul 1995 for more detail).

The frequency of time-based assessments will determine the resolution the study will have. The resolution needed depends on the goal of the study, what is known about the behavior studied, and the theoretical framework of the study.

A variety of time-based assessment schedules are used in EMA. Some administer assessment at fixed intervals. This has been common

in ambulatory monitoring of blood pressure (e.g., every 30 to 45 minutes in Kamarck et al. 1998). Assessment at equal intervals allows the time block to serve as the unit of analysis and supports analyses that require evenly spaced assessments, such as simple autocorrelation analysis and time series analysis. (Daily diaries are a special case discussed below.) Some studies have used somewhat irregular intervals, typically defined by social parameters: For example, Hensley et al. (2003) had subjects complete an asthma diary in the morning and evening of each day. Because the time blocks are vaguely defined, they give subjects considerable discretion in the timing of assessments, and this has the potential to introduce bias. For example, subjects might remember to complete their mood assessments when they are reminded by extreme affect, thus biasing the sample of mood.

An alternative to fixed intervals is a variable schedule, which usually administers assessments at random times in order to achieve a representative sampling of subjects' state. A variation is stratified random sampling, from within strata defined by blocks of time within a day. For example, Affleck and colleagues (1998) assessed subjects once in each predefined time window in the morning, afternoon, or evening, with assessments scheduled at random within each interval. This guarantees that the sample of assessments will evenly sample time across the day and ensures (subject to missed assessments) that each time block includes an assessment, which allows the time block to serve as the unit of analysis.

Setting the frequency of assessments includes considerations of subject burden as well as of how rapidly the target phenomenon is expected to vary. Although assessing subjects 3 to 5 times per day is common, some studies have succeeded with as many as 20 or more assessments per day (Goldstein et al. 1992; Kamarck et al. 1998, 2002, 2005, 2007). Assessment frequency may be varied to collect more assessments at certain times of day if those are of particular interest (Shiffman et al. 2000). Assessments should ideally be sched-

uled throughout the waking day. Some studies have limited assessment to a narrow range of hours (e.g., 10AM to 10PM in Kimhy et al. 2006), which misses early-morning and late-night hours that may encompass important—and substantially different—experiences and behaviors.

When using time-based assessment schedules, especially with variable intervals, EMA studies require some method of signaling subjects when an assessment is scheduled. This is typically accomplished through the use of a device—a beeper, phone, wristwatch, or personal digital assistant (PDA)—programmed to signal the subjects at the appropriate times (see Shiffman 2007 for a discussion of devices).

Thus, although events can be captured at the subject's initiative, continuous phenomena typically have to be sampled using a suitable time-based sampling scheme.

Combination Designs

Different sampling approaches can be fruitfully combined to test particular hypotheses. When the researcher is interested in the circumstances that are associated with a target event, it can be particularly helpful to combine time- and event-based assessments in order to provide a context for interpreting the event data. Data on the situational context of clinically meaningful events, such as panic attacks (Margraf et al. 1987) or binge-eating episodes (Engel et al. 2007) are hard to interpret without comparison to base-rates (Paty et al. 1992). For example, Greeno et al. (2000) found depressed affect among bingers at the time of a binge. However, this is difficult to interpret: Does it mean that negative moods help trigger binges, or just that binge eaters are generally unhappy, even when not bingeing? By comparing affect reported during a binge with affect reported on random occasions outside of binges, Greeno et al. (2000) were able to establish that binges were specifically associated with more distress. This design is modeled after the case-control design, with the events as "cases" and the nonevent

PDA: personal digital assistant (a handheld computer)

data collected by time-based methods constituting the "controls," and is often referred to as a case-crossover design (Maclure & Mittleman 2000).

In EMA designs combining event- and time-based assessments, time-based assessments can also document the antecedents or sequelae of events. For example, Shiffman & Waters (2004) used time-based data to show that ex-smokers were experiencing escalating levels of affective distress in the hours preceding a smoking lapse (the event) and to show that self-efficacy decreased following a lapse, but not after occasions when smokers successfully resisted a temptation to smoke (Shiffman et al. 1997b). Time-based assessments can be used to follow up on the sequelae of a recorded event. For example, to test how quickly migraine medications delivered pain relief and how long the relief lasted, Sheftell and colleagues (2005) had subjects record the onset of migraines and then scheduled a series of assessments as follow-ups.

Combining different schedules of time-based assessments can also be useful. For example, Muraven et al. (2005) assessed social drinking in subjects throughout the day but also scheduled an assessment each morning to ask about hangovers from the previous night's drinking. This illustrates how decisions about design and scheduling of assessments need to be driven by the research questions and by knowledge about the natural history of the target behavior.

Use of Recall in EMA

Although some EMA studies truly focus on the moment, many assessments involve some degree of retrospection. For example, Affleck and colleagues (1998) asked fibromyalgia subjects to report on their pain and fatigue over the past 30 minutes. Event-based diaries often involve some retrospection if the entry is made after the event: For example, studies with the Rochester Interaction Diary (Reis & Wheeler 1991) ask subjects about social interactions they had just concluded. Thus, the

focus on momentary experience is not absolute; it might more liberally be thought of as a focus on recent experience. Of course, some events and experiences are likely to be recalled more accurately, particularly if they are uncommon and/or highly salient (e.g., a major marital spat) or recent. However, since the literature shows that recall can be biased over even a short interval, the researcher needs to carefully consider the use of recall methods and the potential for bias even over short intervals.

Sampling Versus Coverage Strategies

When subjects are assessed intermittently and the focus is on their momentary state, the assessments do not provide complete coverage of their daily experience. That is, even though we may have assessed their mood 12 times in a day, and found them happy each time, it is possible that they experienced moments of misery in between the assessments. A sampling strategy recognizes that only some moments are assessed, but relies on the idea that, over multiple days and multiple assessments, both moments of happiness and moments of misery will be sampled, and the mix will be representative of the subjects' average mood.

In lieu of a sampling strategy, investigators sometimes adopt a coverage strategy, which aims to cover every moment of the day. For some kinds of objective data, continuous measurement (e.g., actigraphy) is possible, enabling true coverage (e.g., Tulen et al. 2001). For self-report measures, continuous assessment is not possible, so investigators try to obtain coverage by having subjects recall and summarize the period between assessments. This is sometimes implemented by asking subjects several times a day to recall and summarize their experience since the last assessment, providing piecemeal coverage of the entire day. For example, Van den Brink et al. (2001) assessed pediatric headache patients four times each day, and at each assessment, asked them to recall and characterize headaches experienced since the previous

assessment. Besides being vulnerable to biases introduced by recall, this strategy requires subjects to accurately maintain a timeline of experience and focus on experience since the last assessment. Research suggests that people are particularly poor at keeping track of time and identifying when events took place (Sudman & Bradburn 1973), suggesting that this approach needs to be treated with caution. Coverage strategies may be useful when the investigator needs an absolute estimate or count of events; otherwise, sampling strategies can yield representative estimates.

Daily Diaries: A Special Case of EMA

Daily (i.e., once-a-day) diaries are particularly popular with researchers and clinicians because of their ease of administration and relatively low subject burden. Within the framework that we have been discussing, daily diaries are fixed-interval assessments with an assessment frequency of once per day, employing a retrospective coverage strategy. Even though daily diaries rely on recall of the entire day and therefore are not momentary, we consider them as falling into the family of EMA designs because they are administered repeatedly and provide a dynamic look at the variables investigated, although with less resolution than that of within-day assessments.

Daily diaries bring some of the benefits of EMA in terms of illuminating process over time, but they also carry with them some of the limitations of retrospective assessments, asking participants to summarize the entire day's experience. This may introduce some bias because moods can vary substantially from one moment to the next, and true recall is likely to stretch the capabilities of autobiographical memory. Moreover, research on autobiographical memory suggests that biases due to heuristic recall strategies operate even over short time frames. The typical scheduling of diary completion at the end of the day may itself become a source of bias, since this tends to be a highly unrepresentative moment, when subjects may be particularly fatigued, for ex-

ample, which could bias their recall of the day's experience. Accurate recall may be more likely when subjects are asked to recall less frequent, more notable and salient events, such as the occurrence of a migraine headache. Comparisons of daily diary data to real-time data will enable researchers to determine whether the former is a reasonable proxy for the latter. Research is needed to shed light on when accurate recall is and is not likely. Thus, daily diaries require a cautious approach.

Besides concern about retrospective bias, daily assessment also provides limited temporal resolution in which to observe behavior; that is, within-day processes may be critically important for some phenomena. This was illustrated in a study by Shiffman & Waters (2004) on the antecedents of smoking lapses. Daily diary data showed no increase in negative affect in the days leading up to a lapse and no association between daily negative affect and lapse risk. However, examination of hour-by-hour affect data revealed a steep increase in negative affect in the hours immediately preceding lapse episodes. Thus, day-level data may miss important sources of dynamic variation that drive behavior. However, when used carefully, especially for highly memorable and slow-moving targets, daily diaries can nevertheless be a tremendous asset for studying certain questions and are often an improvement over relying on global retrospective recall.

USE OF ECOLOGICAL MOMENTARY ASSESSMENT IN TREATMENT

For Assessment

Although we have been addressing the use of EMA for research data, it also seems natural to use EMA for ongoing assessment during treatment. The treatment context includes many features that lend themselves to EMA methods: The client is likely motivated to devote energy to assessment, and there are usually clear target behaviors or experiences on which to focus, dictated by the client's

presenting problem, the nature of the pathology, and/or the clinical formulation of the case. EMA methods lend themselves to idiographic n = 1 analysis; even formal statistical testing can be done by using a single client's data to guide therapy and assess progress (e.g., Delespaul 1995). Properly structured EMA data lend themselves to the sort of microanalysis of process that is often the subject of therapeutic discussion, and that is seen as revealing opportunities for intervention ("Let's figure out what led up to your anxiety attack so we can understand it and think about how to prevent it"). Because change is expected during treatment, ongoing assessment can be informative. EMA could also prove useful in shedding light on processes and mediators of psychotherapy-induced change.

EMA assessments are only beginning to see application in clinical settings. Norton and colleagues (2003) conducted a small outcome study of CBT for binge eating, in which patients were randomized to use EMA-based assessment or not. There were no significant differences in the primary outcome, but there were strong trends favoring the patients with EMA, suggesting that a larger study would have demonstrated reliable differences. Protocols to make optimal use of EMA in treatment have yet to be developed. Widespread use of EMA in treatment will also have to overcome both real and perceived barriers of the technical and financial burden of these methods. We expect that EMA methods will see increasing use in clinical settings over the next decade (see Piasecki et al. 2007 for a review of EMA applied to clinical assessment).

For Intervention

Although using EMA for clinical assessment would be a minor extension of its use in research, applying EMA methods directly to intervention—that is, implementing real-time in-the-moment interventions as subjects go about their daily lives—may open up more radical developments. Bridging the gap between patients' here-and-now experiences and behaviors in the real world and the removed, confined, and limited contact with the clinician in the consulting room has always been a major challenge for psychological treatment. Only a few papers have described momentary interventions, and fewer still have evaluated them.

Newman et al. (2003) reviewed a variety of palm-top-assisted treatments for psychological disorders. Newman et al. (1997) reported that a brief palm-top-assisted momentary intervention for panic disorder was equivalent in efficacy to a longer therapist-administered treatment. The control group in this study also engaged in EMA and may have benefited from that, which highlights the importance of identifying the incremental benefits of EMA intervention versus EMA. Momentary interventions have also been described for eating disorders (Norton et al. 2003) and addictive disorders (Riley et al. 2002). Carter et al. (2007) describe a particularly sophisticated and interactive EMA-based treatment program for smoking (still under evaluation) and present a conceptual foundation of EMA-based treatments.

At this stage, then, EMA-based momentary intervention remains a promising, but only partly proven, idea. However, its potential seems enormous. The idea of momentary treatment, delivering intervention immediately on-the-spot and as needed in real-world settings—a "therapist in your pocket"—holds promise for addressing behavior at crucial moments in the patient's life. EMA-based treatment can provide behavioral guidance or other interventions (e.g., relaxation stimuli) throughout a patient's day as well as just at the moment they are needed. By learning from an individual's history, algorithms could, for example, tailor coping suggestions based on what has worked before for this patient in this situation. Furthermore, by observing the patient over time, predictive algorithms could anticipate and respond to challenges before they gain strength, for example, noting rising stress levels and intervening (or contacting a

counselor) before they result in maladaptive behavior. In-the-moment interventions have only begun to be explored, but they have the potential to revolutionize clinical treatment.

MEASUREMENT CONSIDERATIONS FOR ECOLOGICAL MOMENTARY ASSESSMENT SELF-REPORTS

When self-report assessments are used in EMA, some special considerations are needed. Many questionnaires were not designed for assessing momentary states; the instructions may have a longer recall time frame or no time frame at all. Whether adapting an existing instrument for EMA or creating a new one, it is important to consider whether the item makes sense in the new time frame—i.e., momentary, hourly, etc.—both from the investigator's conceptual perspective and from the subject's perspective. For the latter, cognitive testing of proposed assessments (Willis 2005), as well as pilot testing of assessments in the field, is usually prudent.

An unusual aspect of EMA studies is that subjects will typically encounter each assessment many, many times. This requires that assessments be well tuned: Something that is just a small annoyance when seen once can become a real irritant when encountered five times a day for weeks.

Construction of EMA assessments may also be influenced by the device or medium used. For example, PDAs have limited screen display space, requiring that items and response options be reasonably succinct. When interactive voice response systems are used, the length, complexity, and number of options needs to be severely limited, keeping in mind that subjects must listen to and remember all the response options and their corresponding keypad numbers. Simply moving a questionnaire from paper to a screen-based electronic device should not be problematic: A meta-analysis showed that paper-based and screen-based questionnaires were equivalent (Gwaltney et al. 2007).

Use of programmable electronic devices to administer assessments provides opportunities to change the order in which assessment items are administered. This can be used to present items in random order to counterbalance any effects of item sequencing (for free access to such software, see **http://www.experiencesampling.org/**). This programmability could be used to implement computer-adaptive testing (Chang & Reeve 2005), which provides efficient assessment of a construct with a minimum number of items, by dynamically changing which items are administered to particular subjects based on their responses to prior items.

Although most EMA studies collect structured quantitative data (using structured scales to quantify responses), EMA methods can also be used to gather qualitative data. Diaries can allow write-in responses, and O'Connell et al. (2002) have supplemented PDAs for quantitative assessment with tape recorders for collecting qualitative narratives for later coding. Automatically activated tape recorders have also been used to collect qualitative data (i.e., conversations) without the need for subject intervention (Mehl et al. 2001). Assessments of physiological parameters usually bring with them a set of technical concerns specific to the system being measured and the method used.

Besides keeping in mind the special measurement issues that apply to EMA, it is equally important to recognize that many of the measurement issues that apply to global psychometric assessment also apply to EMA. That is, it is important that assessments be reliable and valid; in EMA, reliability can sometimes be achieved through aggregation across multiple assessments rather than across multiple items within a single assessment. The meaning of EMA assessments has to be carefully considered; Schwarz (2007) has speculated that focusing self-report on immediate experience might shift the individual's focus to very small events at the cost of the "big picture." It is also important to keep in mind that self-report data collected via

EMA methods are still self-report data and are subject to many of the limitations inherent to self-report. Like any self-report data, EMA data can be adversely influenced by the effects of subjects' psychopathology (Kessler et al. 2000) and by deception or self-deception.

METHODOLOGICAL CONSIDERATIONS IN ECOLOGICAL MOMENTARY ASSESSMENT STUDIES

Reactivity

Reactivity is defined as the potential for behavior or experience to be affected by the act of assessing it. Behavioral studies of self-monitoring, in which patients were asked to monitor the behavior they were trying to change, often demonstrated reduction in the problem behavior due to monitoring alone—so much so that self-monitoring came to be considered a part of behavior-change treatment, not just assessment. However, reactivity is not universally observed. In particular, studies showed that reactivity emerged when subjects were trying to change the target behavior and when the behavior was recorded prior to being executed (e.g., recording a meal before eating it)—in other words, under conditions that provide both the motivation and the opportunity to exert control over the behavior (Korotitsch & Nelson-Gray 1999). This suggests that EMA researchers need to be concerned about reactivity under such conditions, although there has not yet been a direct test of this idea.

When such conditions are not present, several studies of EMA find little or no evidence of reactive effects (Cruise et al. 1996; Hufford & Shields 2002; Hufford et al. 2002a,b; Hufford & Shiffman 2002; Litt et al. 1998; Stone et al. 1998). In the most controlled study of reactivity to EMA, Stone and colleagues (2003a) assigned subjects with pain syndromes to complete either no EMA monitoring or sampling of their pain using electronic diaries 3, 6, or 12 times daily. No evidence was found that the pain ratings were systematically reactive to the EMA monitoring.

Besides traditional reactivity, one might also be concerned that the burden of monitoring, including being interrupted by "beeping" throughout the day, might cause distress that would bias the data. Again, the evidence suggests this is not typical. The Stone et al. (2003a) study found no trend over time in pain ratings and no systematic effect on pain ratings of increased prompting, despite a deliberately aggressive prompting and assessment schedule.

Given the earlier literature suggesting that reactivity can occur, the literature does not completely resolve when reactive effects might or might not be observed. EMA researchers should be alert to the potential for reactivity while recognizing that little evidence has been found to support the concern that EMA engenders significant reactivity.

Compliance

While EMA methods promise to capture data closer in time to events and symptoms of interest, they also place on subjects the burden of making timely recordings as they go about their daily lives. Missing assessments have the potential to bias the obtained sample of behavior and experience, especially if the missing data are nonrandom. As a result, EMA studies place a premium on obtaining high levels of subject compliance with the assessment protocol.

The methods used to collect diary data can affect the investigator's ability to obtain and document compliance. Studies using paper diaries often report that about 90% of diary cards are returned (Hufford & Shields 2002). However, timely compliance is often inferred from the date and time subjects have recorded on the diary card. This opens the door for subjects to hoard and backfill diaries (that is, to complete entries, perhaps en masse, after the fact) while seeming compliant.

Concern about hoarding and backfilling is heightened by the findings from studies in which subjects completed written diaries while also using electronically instrumented devices (e.g., inhalers, blood glucose monitors) that independently tracked the actual time and date of events, such as medication taking or glucose monitoring. Across eight studies (Hufford 2007), objective data from the devices indicated that compliance had actually been much lower than was indicated by the written diaries: Many of the written entries had been falsified. Moreover, the written entries were systematically biased, for example, showing better glycemic control than had actually been achieved (Mazze et al. 1984). The written diary records understated problems of glucose control and accentuated the concern about backfilling and faked compliance. These concerns were further accentuated by reports from paper diary studies in which high proportions of subjects later admitted to backfilling entries, even when they had been electronically prompted at the appropriate times and had recorded those times on their diary cards (Litt et al. 1998, Moghaddam & Ferguson 2007).

To assess true compliance with written diaries, Stone et al. (2002) used an instrumented paper diary (IPD), which was covertly instrumented with photosensors that recorded the opening of the diary booklet—a prerequisite to completing diary cards. Chronic pain subjects were assigned to use either the IPD or a PDA-based electronic diary to record their pain at three scheduled times daily for three weeks. Subjects were trained on the diary, and steps were taken to encourage motivation and compliance. The IPD allowed comparison of reported compliance—based on the time and date handwritten on returned diary cards—and actual compliance—based on the electronic records. Reported compliance was 90%, as in many other paper diary studies. However, actual compliance was estimated at 11%. In other words, almost 90% of the submitted diary cards had been falsified. Evidence indicated that the diaries had been hoarded

for days and backfilled. More disturbing still, many subjects also showed evidence of filling their paper diary cards in advance of the stated time (Stone et al. 2003b), which obviously renders the data invalid. Broderick et al. (2003) subsequently showed that adding an audible reminder to complete the diary boosted compliance only modestly.

The other 40 pain sufferers in the study by Stone et al. (Stone et al. 2002, 2003a,b; Stone & Shiffman 2002) had been given an electronic diary, which prompted for assessments and did not allow entries outside the designated time windows. These subjects actually did complete 94% of their assigned assessments on time.

These compliance findings by Stone and colleagues have stirred debate among diary researchers (Green et al. 2006, Tennen et al. 2006), with critics noting, for example, that various procedures (e.g., having subjects return diaries daily by mail, so that postmarks provide some degree of time-stamping) might improve timely compliance with written diaries. In any case, the Stone et al. study (Stone et al. 2002, 2003a,b; Stone & Shiffman 1994, 2002), along with the prior literature documenting falsified paper diary entries (Hufford 2007), raises critical questions about timely compliance. Paper-and-pencil diaries are vulnerable to being falsified, and both forward- and backfilling are possible and potentially frequent, making verification of timely compliance both essential and challenging. Thus, the burden of proof falls to the EMA investigator to establish, by whatever means, the timeliness of recording.

The most common way to ensure and document timely compliance is the use of electronic diaries, which tag each record with time of entry and thus allow for detailed analysis of compliance, particularly with interval- or signal-contingent recordings. Many electronic diary studies document compliance rates around 90% (Hufford & Shields 2002). However, good compliance is not universal even with electronic diaries; rates as low as 50% have been reported (e.g., Jamison

et al. 2001), suggesting that the mere use of electronic diary technology is not, by itself, a panacea. A variety of suggestions of ways to enhance compliance with electronic diary studies has been presented elsewhere (Hufford & Shiffman 2003). As noted above, compliance with event-oriented protocols is difficult to verify.

SPECIAL POPULATIONS

Because of the burdens they place on subjects, EMA methods often raise concerns about application to special populations. These concerns are often magnified when EMA studies employ electronic devices, such as PDAs or cell phones, for data collection. A common concern is that elderly subjects will be unable or unwilling to operate high-tech devices. However, data on subject preference for electronic versus more traditional paper diaries do not support this assumption. In a meta-analysis of patient preference, older subjects are actually less likely to prefer paper diaries (in Shiffman et al. 2007). Similarly, although there have been concerns about the use of such technology in low-education, low-socioeconomic-status subjects, Finkelstein et al. (2000) demonstrated good success with just such a sample.

Different concerns are sometimes raised about the use of diaries with children. Several studies have demonstrated that children are able to use electronic diaries, and good compliance and results have been seen with children as young as 7 (see Van den Brink et al. 2001). Of course, the use of EMA methods does not obviate, and may increase, some more general issues about assessment of children, such as comprehension of the questionnaires used.

Concerns also arise about whether some clinical populations might be unable to perform in EMA studies precisely because of their psychopathology. In this regard, it is notable that several studies have been successfully carried out among schizophrenics (Delespaul 1995, Myin-Germeys et al. 2001), and good

compliance using cell phones for EMA reporting has been reported in a sample of homeless crack addicts (Freedman et al. 2006). These examples suggest that, with appropriate sensitivity, EMA methods and related technology may be useful in a wide array of patient populations and that investigators should not make decisions about feasibility without testing their assumptions.

A subtler concern that is not addressed by the performance of special populations in EMA studies is that subjects who volunteer for and complete demanding EMA studies may not be representative. There have been no reports of particular recruitment difficulties or notable deviations in EMA study samples, but this has not been systematically evaluated. It seems likely that personal and environmental factors may discourage some subjects' participation in EMA studies. Occupational demands may preclude availability for momentary assessment (consider surgeons and assembly line workers, although see Goldstein et al. 1992). Also, some work environments may be so noisy as to preclude hearing prompts or using a phone. Motor impairments, impaired vision or hearing, and illiteracy may preclude participation in EMA (and in many non-EMA) studies. Finally, subjects who are unfamiliar with or fearful of technology may be put off by EMA studies using high-tech devices. The limits of EMA and their impact on studies should be continually evaluated.

ANALYSIS OF EMA DATA

EMA data usually consist of a large number of observations from each subject, with the number and timing of observations often varying between subjects. EMA data do not lend themselves to the basic approaches most clinical psychologists learned in graduate-school statistics classes, which require independent observations. Nor do they fit easily into traditional repeated-measures designs, which typically require the same number of observations for each person and no missing data. However,

a variety of methods—essentially extensions of the regression framework most psychologists are familiar with—can handle such data. Known variously as random effects models, multilevel or hierarchical analysis, and generalized estimating equations, these methods implement regression models while accounting for the interrelationships of observations within subjects. Extensions of these methods go beyond linear regression to cover logistic regression and survival models, among others. Detailed exposition about these methods is beyond the scope of this review; the reader is referred to Schwartz & Stone (1998, 2007) for detailed and practical guidance on applications of these models to EMA data. The bottom line is that EMA researchers can use easily available software to apply the conceptual models they are familiar with from more traditional between-subjects analyses to EMA data. Besides these relatively familiar statistical models, a variety of novel approaches can fruitfully be applied to EMA data (Walls & Schafer 2006).

PRACTICAL ISSUES IN ECOLOGICAL MOMENTARY ASSESSMENT

The implementation of EMA places unique demands on both researchers and subjects. A detailed discussion of the practical aspects of implementing EMA is beyond the scope of this review. The interested reader is referred to Hufford (2007). Here we touch only briefly on considerations of hardware and software, and management of subjects and data.

EMA Hardware and Software

An investigator planning an EMA study faces many practical issues and decisions. One key decision is what type of mechanism to use to capture data. Although EMA is most closely associated with the use of palm-top computers (Hufford & Shields 2002), the method is hardware agnostic and encompasses a range of high- and low-tech devices, ranging from

paper and pencil to futuristic wireless sensors (Intille 2007). The key factor to consider in selecting a hardware platform is the support provided for key functions, including the ability to prompt subjects to complete assessments, manage prompting schedules (including randomization and contingencies), present assessment content to subjects, manage assessment logic such as branching, accept and store subjects' self-report data, time-stamp entries to avoid faked compliance, and/or directly collect physical or physiological data (see Shiffman 2007). Investigators should also consider the suitability of a particular device or medium to their subject population: For example, a small screen or a telephone system may be less suitable for an elderly population with visual or hearing impairments, respectively.

The cost of devices and of software development is often an additional consideration. Investigators should avoid being overly influenced by initial start-up costs and also should consider downstream costs, such as the cost of cell phone service or of data entry from paper diaries. Unfortunately, high fixed costs of technological solutions can make it very hard to initiate small studies or pilot studies.

EMA approaches using electronic devices require software for basic system functionality and for the specifics of the study protocol. The success of such studies depends on correct and robust performance of software, which can often be complex. Both freeware and commercial versions of EMA software are now readily available. Le et al. (2006) have reviewed a number of freely available programs, and a number of commercial vendors also provide programming and support services for EMA studies. These base systems have to be configured or customized to the needs of a particular study. Researchers should carefully consider capabilities and limitations of various software systems. For example, some systems can only accommodate signaling during specified periods of the day, which can bias the sample by, for example, omitting early-morning and late-night events, which may be highly

salient. Some programs may support event- or signal-driven assessments, but not the combination of the two. In light of the availability of software from multiple sources, investigators are well advised to use existing software or services.

Other Practical Considerations

Investigators should also consider procedures and resources related to managing study subjects. Training of subjects on the protocol, the assessment, and any device used is an essential aspect of EMA studies. Data management is an additional consideration in EMA studies. EMA studies can produce voluminous datasets with hundreds of thousands of observations, possibly of diverse kinds (e.g., momentary, daily, events). Additional complications include the fact that the records are ordered in time and the temporal relationships need to be maintained, and there may be nondata records that do not represent subject assessments (e.g., records of missed assessment prompts) that also have to be maintained within the database. Data management facilities for maintaining these databases and producing analytical datasets are an essential resource for EMA studies and should be planned into the study.

CONCLUSION

EMA studies help us appreciate the importance of change and context in everyday behavior and experience. The dynamic influence of context is demonstrated not only for constructs that we might expect to be variable, such as affect, craving, and pain, but also for many constructs that are typically conceptualized as stable trait-like characteristics, such as extraversion (Fleeson 2001), self-esteem (Kernis 2005), coping style (Schwartz et al. 1999), and self-efficacy (Gwaltney et al. 2005). EMA's microscopic focus and temporal resolution is beginning to help us see the underlying dynamics behind even trait-like constructs. EMA studies suggest that behavior and experience are much more dynamic, and much more influenced by its immediate context, than we usually envision. Even individual differences can sometimes be best understood through analysis of within-person variation. For example, differences in how persons react to situational stimuli, such as stressors, are related to personality (Feldman-Barrett 1997), psychopathology (Myin-Germeys et al. 2001), and clinical outcomes (Kamarck et al. 2007), and within-person variation in response may sometimes be as important as the mean level (Gwaltney et al. 2005, Kernis 2005). Despite their limitations, EMA methods promise to enhance our understanding of the dynamic interactions between individuals and their environments.

Clinical psychology—and psychology in general—is prone to overestimate the role of stable traits in determining behavior and to underestimate the influence of the local setting on behavior (Mischel 2004). Our behavior in daily life is influenced not just by our predispositions, but also by where we are and whom we are with, by how we are feeling and what situation we are in, by what has recently happened and what we have done or felt in the minutes and hours preceding the present moment. One context can elicit dysfunctional behavior while another elicits a healthier response; one can fuel psychopathology or set us on a course toward improvement. Understanding life as it is lived, up close, will help us better understand both health and pathology and help us see where there are opportunities to intervene on the side of health. This requires methods that examine behavior at the appropriate level of granularity: "Research methods that examine this landscape from 10,000 feet cannot shed light on how this landscape is shaped at ground level" (Shiffman 2005, p. 1743). EMA methods provide an important tool to help clinical psychology explore the dynamic nature of behavior, thought, and feeling as they unfold over time.

SUMMARY POINTS

1. Autobiographical memory processes can introduce bias into retrospective self-reports, which form the bulk of clinical assessments.

2. EMA is a collection of methods for obtaining repeated real-time assessments of subjects' behavior and experience in their natural environments, to minimize recall bias, maximize ecological validity, and document variation over time.

3. EMA encompasses a variety of diary approaches and technologies used to collect data on a schedule (e.g., daily diaries or assessments scheduled at random) or in response to clinical events (e.g., symptom episodes or behaviors).

4. The collection of frequent assessments makes EMA studies well suited to study microprocesses—i.e., how behavior and experience vary over time and across changing contexts.

5. Although sometimes retrospective self-reports correspond well with those based on aggregated EMA data, recall of behavior and experience over time does not correspond well to real-time EMA assessments.

6. The application of EMA methods for assessment and intervention in treatment is promising but is in its infancy.

7. EMA methods have great potential to advance the science and practice of clinical psychology by providing more valid and more detailed data about real-world behavior and experience.

FUTURE ISSUES

1. Further prospective longitudinal analyses of EMA data are necessary to shed light on theoretically relevant microprocesses in studies of psychopathology and treatment.

2. Further improvements in technological tools for EMA studies would enhance ease of use, reduce cost, and expand capabilities.

3. An improved understanding is needed of the relationship between recall and real-time EMA data, the circumstances under which recall may be accurate, and the processes by which persons develop retrospective accounts of their experience.

4. Models are needed for application of EMA methods to clinical assessment and intervention.

ACKNOWLEDGMENTS

DISCLOSURE STATEMENT

LITERATURE CITED

Affleck G, Tennen H, Urrows S, Higgens P, Abeles M, et al. 1998. Fibromyalgia and women's pursuit of personal goals: a daily process analysis. *Health Psychol.* 17:40–47

Barker R. 1978. *Habitats, Environments and Human Behavior: Studies in the Ecological Psychology and Ecobehavioral Science of the Midwest Psychological Field Station: 1947–1972.* San Francisco, CA: Jossey-Bass

Beal DJ, Weiss HM. 2003. Methods of Ecological Momentary Assessment in organizational research. *Organ. Res. Methods* 6:440–64

Bolger N, Davis A, Rafaeli E. 2003. Diary methods: capturing life as it is lived. *Annu. Rev. Psychol.* 54:579–616

Bolger N, Laurenceau J-P. 2005. Using diary methods to study marital and family processes. *J. Family Psychol.* 19:86–97

Bradburn N, Rips L, Shevell S. 1987. Answering autobiographical questions: the impact of memory and inference on surveys. *Science* 236:157–61

Broderick JE, Schwartz JE, Shiffman S, Hufford MR, Stone AA. 2003. Signaling does not adequately improve diary compliance. *Ann. Behav. Med.* 26:193–248

Broderick JE, Schwartz JE, Stone AA. 2006. *Context (pain and affect) influences recall pain ratings.* Presented at Annu. Meet. Am. Pain Soc., San Antonio, TX

Byerly M, Fisher R, Whatley K, Holland R, Varghese F, et al. 2005. A comparison of electronic monitoring vs clinician rating of antipsychotic adherence in outpatients with schizophrenia. *Psychiatry Res.* 133:129–33

Carney MA, Tennen H, Affleck G, del Boca FK, Kranzler HR. 1998. Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real-time "electronic interviews." *J. Stud. Alcohol* 59:447–54

Carter BL, Day SX, Cinciripini PM, Wetter DW. 2007. Momentary health interventions: Where are we and where are we going? See Stone et al. 2007b, pp. 289–307

Chang C, Reeve BB. 2005. Item response theory and its applications to patient-reported outcomes measurement. *Eval. Health Prof.* 28:264–82

Clark DM, Teasdale JD. 1982. Diurnal variation in clinical depression and accessibility of memories of positive and negative experiences. *J. Abnorm. Psychol.* 91:87–95

Cramer JA, Mattson RH, Prevey ML, Scheyer RD, Ouellette VL. 1989. How often is medication taken as prescribed: a novel assessment technique. *J. Am. Med. Assoc.* 261:3273–77

Cruise CE, Broderick JE, Porter L, Kaell A. 1996. Reactive effects of diary self-assessment in chronic pain patients. *Pain* 67:253–58

Curry S, Marlatt GA, Gordon JR. 1987. Abstinence violation effect: validation of an attributional construct with smoking cessation. *J. Consult. Clin. Psychol.* 55:145–49

Delespaul P. 1995. *Assessing Schizophrenia in Daily Life: The Experience Sampling Method.* Maastricht, The Netherlands: Maastricht Univ. Press

DeVries M, ed. 1992. *The Experience of Psychopathology: Investigating Mental Disorders in Their Natural Settings.* Cambridge, UK: Cambridge Univ. Press. 429 pp.

Eich E, Reeves J, Jaeger B, Graff-Radford S. 1985. Memory for pain: relation between past and present pain intensity. *Pain* 223:375–79

Engel SG, Boseck J, Crosby RD, Wonderlich SA, Mitchell JE, et al. 2007. The relationship of momentary anger and impulsivity to bulimic behavior. *Behav. Res. Ther.* 45:437–47

Fahrenberg J, Myrtek M. 2001. *Progress in Ambulatory Assessment: Computer-Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies.* Seattle, WA: Hogrefe & Huber

Summarizes biases related to autobiographical memory, and how they can bias research data.

Feldman-Barrett L. 1997. The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personal. Soc. Psychol. Bull.* 23:1100–10

Feldman-Barrett L, Russell JA. 1998. Independence and bipolarity in the structure of current affect. *J. Personal. Soc. Psychol.* 74:967–84

Finkelstein J, Cabrera MR, Hripcsak G. 2000. Internet-based home asthma telemonitoring. *Chest* 117:148–55

Fleeson W. 2001. Toward a structure- and process-integrated view of personality: traits as density distributions of states. *J. Personal. Soc. Psychol.* 80:1011–27

Freedman M, Lester K, McNamara C, Milby J, Schumacher J. 2006. Cell phones for Ecological Momentary Assessment with cocaine-addicted homeless patients in treatment. *J. Subst. Abuse Treat.* 30:105–11

Goldstein I, Jamner L, Shapiro D. 1992. Ambulatory blood pressure and heart rate in healthy male paramedics during a workday and a nonworkday. *Health Psychol.* 11:48–54

Green AS, Rafaeli E, Bolger N, Shrout PE, Reis HT. 2006. Paper or plastic? Data equivalence in paper and electronic diaries. *Psychol. Methods* 11:87–105

Greeno CG, Wing R, Shiffman S. 2000. Binge antecedents in obese women with and without Binge Eating Disorder. *J. Consult. Clin. Psychol.* 68:95–102

Gwaltney CJ, Shields AS, Shiffman S. 2007. Equivalence of electronic and paper-and-pencil administration of patient reported outcome measures: a meta-analytic review. *Value Health*. In press

Gwaltney CJ, Shiffman S, Sayette MA. 2005. Situational correlates of abstinence self-efficacy. *J. Abnorm. Psychol.* 114:649–60

Hektner JM, Schmidt JA, Csikszentmihalyi M. 2007. *Experience Sampling Method: Measuring the Quality of Everyday Life*. Thousand Oaks, CA: Sage

Hensley MJ, Chalmers A, Clover K, Gibson PG, Toneguzzi R, Lewis PR. 2003. Symptoms of asthma: comparison of a parent-completed retrospective questionnaire with a prospective daily symptom diary. *Pediatr. Pulmonol.* 36:509–13

Homma Y, Ando T, Yoshida M, Kageyama S, Takei M, et al. 2002. Voiding and incontinence frequencies: variability of diary data and required diary length. *Neurourol. Urodyn.* 21:204–9

Houtveen JH, Oei NYL. 2007. Recall bias in reporting medically unexplained symptoms comes from semantic memory. *J. Psychosom. Res.* 62:277–82

Hufford MR. 2007. Special methodological challenges and opportunities in Ecological Momentary Assessment. See Stone et al. 2007b, pp. 54–75

Hufford MR, Shields AL. 2002. Electronic diaries: an examination of applications and what works in the field. *Appl. Clin. Trials* 11:46–56

Hufford MR, Shields AL, Shiffman S, Paty J, Balabanis M. 2002a. Reactivity to Ecological Momentary Assessment: an example using undergraduate problem drinkers. *Psychol. Addict. Behav.* 16:205–11

Hufford MR, Shiffman SS. 2002. Methodological issues affecting the value of patient-reported outcomes data. *Exp. Rev. Pharmacoecon. Health Outcomes* 2:119–28

Hufford MR, Shiffman S. 2003. Patient-reported outcomes: assessment methods. *Dis. Manag. Health Outcomes* 11:77–86

Hufford MR, Stone AA, Shiffman S, Schwartz JE, Broderick JE. 2002b. Paper vs. electronic diaries: compliance and subject evaluations. *Appl. Clin. Trials* 11(8):38–43

Intille S. 2007. Technological innovations enabling automatic, context-sensitive Ecological Momentary Assessment. See Stone et al. 2007b, pp. 308–37

Describes the development of the Experience Sampling Method and reviews findings from studies applying it to a variety of domains.

Jamison RN, Raymond SA, Levine JG, Slawsby EA, Nedeljkovic SS, Katz NP. 2001. Electronic diaries for monitoring chronic pain: one-year validation study. *Pain* 91:277–85

Jonasson G, Carlsen K, Sodal A, Jonasson C, Mowinckel P. 1999. Patient compliance in a clinical trial with inhaled budesonide in children with mild asthma. *Eur. Respir. J.* 14:150–54

Kamarck TW, Janicki DL, Shiffman S, Polk DE, Muldoon MF, et al. 2002. Psychosocial demands and ambulatory blood pressure: a field assessment approach. *Physiol. Behav.* 77:699–704

Kamarck TW, Muldoon M, Shiffman S, Sutton-Tyrrell K. 2007. Experiences of demand and control during daily life are predictors of carotid atherosclerotic progression among healthy men. *Health Psychol.* 26:324–32

Kamarck TW, Schwartz J, Shiffman S, Muldoon M, Sutton-Tyrrell K, Janicki D. 2005. Psychosocial stress and cardiovascular risk: What is the role of daily experience? *J. Pers.* 73:1749–74

Kamarck TW, Shiffman S, Smithline L, Goodie JL, Paty J, et al. 1998. Effects of task strain, social conflict, and emotional activation on ambulatory cardiovascular activity: daily life consequences of recurring stress in a multiethnic adult sample. *Health Psychol.* 17:17–29

Kernis MH. 2005. Measuring self-esteem in context: the importance of stability of self-esteem in psychological functioning. *J. Personal.* 73:1–37

Kessler R, Wittchen H-U, Abelson J, Zhao S. 2000. Methodological issues in assessing psychiatric disorders with self-reports. See Stone et al. 2000, pp. 229–55

Kihlstrom J, Eich E, Sandbrand D, Tobias B. 2000. Emotion and memory: implications for self-report. See Stone et al. 2000, pp. 81–100

Kimhy D, Delespaul P, Corcoran C, Ahn H, Yale S, Malaspina D. 2006. Computerized experience sampling method (ESMc): assessing feasibility and validity among individuals with schizophrenia. *J. Psychiatr. Res.* 40:221–30

Kop W, Verdino R, Gottdiener J, O'Leary S, Bairey Merz C, Krantz D. 2001. Changes in heart rate and heart rate variability before ambulatory ischemic events. *J. Am. Coll. Cardiol.* 38:742–49

Korotitsch WJ, Nelson-Gray RO. 1999. An overview of self-monitoring research in assessment and treatment. *Psychol. Assess.* 2:415–25

Larson R, Richards MH. 1994. *Divergent Realities: The Emotional Lives of Mothers, Fathers and Adolescents*. New York: Basic Books

Le B, Hat NC, Beal DJ. 2006. Pocket-sized psychology studies: exploring daily diary software for PalmPilots. *Behav. Res. Methods* 38:325–32

Litt MD, Cooney NL, Morse P. 1998. Ecological Momentary Assessment (EMA) with treated alcoholics: methodological problems and potential solutions. *Health Psychol.* 17:48–52

Litt MD, Cooney NL, Morse P. 2000. Reactivity to alcohol-related stimuli in the laboratory and in the field: predictors of craving in treated alcoholics. *Addiction* 95:889–900

Maclure M, Mittleman MA. 2000. Should we use a case-crossover design? *Annu. Rev. Public Health* 21:193–221

Margraf J, Taylor CB, Ehlers A, Roth WT. 1987. Panic attacks in the natural environment. *J. Nerv. Ment. Dis.* 175:558–65

Mazze RS, Shamoon H, Pasmantier R, Lucido D, Murphy J, et al. 1984. Reliability of blood glucose monitoring by subjects with diabetes mellitus. *Am. J. Med.* 77:211–17

McCarberg BH. 2007. The treatment of breakthrough pain. *Pain Med.* 8:S8–13

McCarthy DE, Piasecki TM, Fiore MC, Baker TB. 2006. Life before and after quitting smoking: an electronic diary study. *J. Abnorm. Psychol.* 115:454–66

McFarland C, Ross M, DeCourville N. 1989. Women's theories of menstruation and biases in recall of menstrual symptoms. *J. Personal. Soc. Psychol.* 57:522–31

McKenzie S, Paty J, Grogan D, Rosano M, Curry L, et al. 2004. Trial power: how eDiary data can enhance the power and reduce the sample size in clinical trials. *Appl. Clin. Trials* 13:54–68

Mehl MR, Pennebaker JW, Crow DM, Dabbs J, Price JH. 2001. The Electronically Activated Recorder (EAR): a device for sampling naturalistic daily activities and conversations. *Behav. Res. Methods Instrum. Comput.* 33:517–23

Mischel W. 2004. Toward an integrative science of the person. *Annu. Rev. Psychol.* 55:1–22

Moghaddam NG, Ferguson E. 2007. Smoking, mood regulation, and personality: an event-sampling exploration of potential models and moderation. *J. Personal.* 75:451–78

Moskowitz D, Young SN. 2006. Ecological Momentary Assessment: what it is and why it is a method of the future in clinical psychopharmacology. *J. Psychiatry Neurosci.* 31:13–20

Muraven M, Collins RL, Morshimer ET, Shiffman S, Paty JA. 2005. The morning after: limit violations and the self-regulation of alcohol consumption. *Psychol. Addict. Behav.* 19:253–62

Myin-Germeys I, van Os J, Schwartz JE, Stone AA, Delespaul PA. 2001. Emotional reactivity to daily life stress in psychosis. *Arch. Gen. Psychiatry* 58:1137–44

Newman MG, Erickson T, Przeworski A, Dzus E. 2003. Self-help and minimal-contact therapies for anxiety disorders: Is human contact necessary for therapeutic efficacy? *J. Clin. Psychol.* 59:251–74

Newman MG, Kenardy J, Herman S, Taylor CB. 1997. Comparison of cognitive behavioral treatment of panic disorder with computer-assisted brief cognitive behavior treatment. *J. Consult. Clin. Psychol.* 65:178–83

Niere K, Jerak A. 2004. Measurement of headache frequency, intensity and duration: comparison of patient report by questionnaire and headache diary. *Physiother. Res. Int.* 9:149–56

Nived O, Sturfelt G, Eckernäs S-Å, Singer P. 1994. A comparison of 6 months' compliance of patients with rheumatoid arthritis treated with tenoxicam and naproxen: use of patient computer data to assess response to treatment. *J. Rheumatol.* 21:1537–41

Norton M, Wonderlich SA, Myers T, Mitchell JE, Crosby RD. 2003. The use of palmtop computers in the treatment of bulimia nervosa. *Eur. Eat. Disord.* 11:231–42

O'Connell KA, Gerkovich MM, Bott M, Cook MR, Shiffman S. 2002. The effect of anticipatory strategies on the first day of smoking cessation. *Psychol. Addict. Behav.* 16:150–56

Oishi S, Sullivan HW. 2006. The predictive value of daily vs. retrospective well-being judgments in relationship stability. *J. Exp. Soc. Psychol.* 42:460–70

Paty JA, Kassel JD, Shiffman S. 1992. Assessing stimulus control of smoking: the importance of base rates. In *The Experience of Psychopathology*, ed. M DeVries, pp. 347–52. Cambridge, UK: Cambridge Univ. Press

Pearson J. 2004. *Better, faster and cheaper diary trials? Evaluating the potential scientific benefits of e-diary technology in a chronic primary insomnia trial*. Presented at Annu. Meet. DIA, Arlington, VA

Perrine MW, Mundt JC, Searles JS, Lester LS. 1995. Validation of daily self-reported alcohol consumption using Interactive Voice Response (IVR) technology. *J. Stud. Alcohol* 56:487–90

Piasecki TM, Hufford MR, Solhan M, Trull TJ. 2007. Assessing clients in their natural environments with electronic diaries: rationale, benefits, limitations, and barriers. *Psychol. Assess.* 19:25–43

Redelmeier DA, Katz J, Kahneman D. 2003. Memories of colonoscopy: a randomized trial. *Pain* 104:187–94

Reis HT, Wheeler L. 1991. Studying social interaction with the Rochester Interaction Record. In *Advances in Experimental Social Psychology*, ed. MP Zanna, pp. 270–318. San Diego, CA: Academic

Riley WT, Jerome A, Behar A, Weil J. 2002. Computer and manual self-help behavioral strategies for smoking reduction: initial feasibility and one-year follow-up. *Nicotine Tobacco Res.* 4:S183–88

Ross M. 1989. Relation of implicit theories to the construction of personal histories. *Psychol. Rev.* 96:341–57

Saito M, Kumano H, Yoshiuchi K, Kokubo K, Yamamoto Y, et al. 2005. Symptom profile of multiple chemical sensitivity in actual life. *Psychosom. Med.* 67:318–25

Schlundt DG, Johnson WG, Jarrell MP. 1985. A naturalistic functional analysis of eating behavior in bulimia and obesity. *Adv. Behav. Res. Ther.* 7:149–62

Schwartz JE, Neale J, Marco C, Shiffman S, Stone AA. 1999. Does trait coping exist? A momentary assessment approach to the evaluation of traits. *J. Personal. Soc. Psychol.* 77:360–69

Schwartz JE, Stone AA. 1998. Data analysis for EMA studies. *Health Psychol.* 17:6–16

Schwartz JE, Stone AA. 2007. Analysis of real-time momentary data: a practical guide. See Stone et al. 2007b, pp. 76–113

Schwarz N. 2007. Retrospective and concurrent self-reports: the rationale for real-time data capture. See Stone et al. 2007b, pp. 11–26

Scollon CN, Kim-Prieto C, Diener E. 2003. Experience sampling: promises and pitfalls, strengths and weaknesses. *J. Happiness Stud.* 4:5–34

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, Jamner LD, Davydov DM, James P. 2002. Situations and moods associated with smoking in everyday life. *Psychol. Addict. Behav.* 16:342–45

Sheftell FD, Dahlöf CG, Brandes JL, Agosti R, Jones MW, Barrett PS. 2005. Two replicate randomized, double-blind, placebo-controlled trials of the time to onset of pain relief in the acute treatment of migraine with a fast-disintegrating/rapid-release formulation of sumatriptan tablets. *Clin. Ther.* 27:407–17

Shiffman S. 1993. Assessing smoking patterns and motives. *J. Consult. Clin. Psychol.* 61:732–42

Shiffman S. 2005. Dynamic influences on smoking relapse process. *J. Personal.* 73:1715–48

Shiffman S. 2007. Designing protocols for Ecological Momentary Assessment. See Stone et al. 2007b, pp. 27–53

Shiffman S, Balabanis MH, Gwaltney CJ, Paty JA, Gnys M, et al. 2007. Prediction of lapse from associations between smoking and situational antecedents assessed by Ecological Momentary Assessment. *Drug Alcohol Depend.* 91:159–68

Shiffman S, Elash CA, Paton S, Gwaltney CJ, Paty JA, et al. 2000. Comparative efficacy of 24-hour and 16-hour transdermal nicotine patch for relief of craving and withdrawal. *Addiction* 95:1185–95

Shiffman S, Engberg J, Paty JA, Perz W, Gnys M, et al. 1997a. A day at a time: predicting smoking lapse from daily urge. *J. Abnorm. Psychol.* 106:104–16

Shiffman S, Ferguson SG, Gwaltney CJ, Balabanis MH, Shadel WG. 2006. Reduction of abstinence-induced withdrawal and craving using nicotine replacement therapy. *Psychopharmacology* 184:637–44

Shiffman S, Gwaltney CJ, Balabanis M, Liu KS, Paty JA, et al. 2002. Immediate antecedents of cigarette smoking: an analysis from Ecological Momentary Assessment. *J. Abnorm. Psychol.* 111:531–45

Shiffman S, Gwaltney CJ, Shields AL. 2007. *Validation of ePRO measures and using ePRO to enhance validity*. Presented at Annu. Meet. New Clin. Drug Eval. Unit, 47th, Boca Raton, FL

Shiffman S, Hickcox M, Paty JA, Gnys M, Kassel JD, Richards T. 1996a. Progression from a smoking lapse to relapse: prediction from abstinence violation effects and nicotine dependence. *J. Consult. Clin. Psychol.* 64:993

Shiffman S, Hickcox M, Paty JA, Gnys M, Kassel JD, Richards T. 1997b. The abstinence violation effect following smoking lapses and temptations. *Cogn. Ther. Res.* 21:497–523

Shiffman S, Hufford M, Hickcox M, Paty JA, Gnys M, Kassel JD. 1997c. Remember that? A comparison of real-time vs. retrospective recall of smoking lapses. *J. Consult. Clin. Psychol.* 65:292–300

Shiffman S, Paty JA. 2003. *Using an electronic diary to capture frequent events*. Presented at From Quality of Life to Patient Outcomes Assess.: Res. Agenda for a Paradigm Shift, Drug Inf. Assoc., Baltimore, MD

Shiffman S, Paty JA, Gnys M, Kassel JD, Hickcox M. 1996b. First lapses to smoking: within-subjects analyses of real-time reports. *J. Consult. Clin. Psychol.* 64:366–79

Shiffman S, Waters AJ. 2004. Negative affect and smoking lapses: a prospective analysis. *J. Consult. Clin. Psychol.* 72:192–201

Shiffman S, Waters AJ, Hickcox M. 2004. The Nicotine Dependence Syndrome Scale: a multidimensional measure of nicotine dependence. *Nicotine Tob. Res.* 6:327–48

Shrier LA, Shih M-C, Beardslee WR. 2005. Affect and sexual behavior in adolescents: a review of the literature and comparison of momentary sampling with diary and retrospective self-report methods of measurement. *Pediatrics* 115:573–81

Stone AA, Bachrach CA, Jobe JB, Kurtzman HS, Cain VS, eds. 2000. *The Science of Self-Report: Implications for Research and Practice*. Mahwah, NJ: Erlbaum

Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. 2003a. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain* 104:343–51

Stone AA, Schwartz JE, Broderick JE, Shiffman S. 2005. Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. *Personal. Soc. Psychol. Bull.* 31:1340–46

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 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

Stone AA, Shiffman S, Atienza A, Nebeling L. 2007a. Introduction to Ecological Momentary Assessment (EMA). See Stone et al. 2007b, pp. xi–xiv

Stone AA, Shiffman S, Atienza A, Nebeling L, eds. 2007b. *The Science of Real-Time Data Capture: Self-Reports in Health Research*. New York: Oxford Univ. Press

Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. 2002. Patient noncompliance with paper diaries. *Br. Med. J.* 324:1193–94

Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. 2003b. Patient compliance with paper and electronic diaries. *Controlled Clin. Trials* 24:182–99

Sudman S, Bradburn NM. 1973. Effects of time and memory factors on response in surveys. *J. Am. Stat. Assoc.* 68:805–15

Initially stated EMA principles, describing the rationale for and characteristics of EMA.

Lays out rationale and foundation of EMA methods, including design, practicalities, and data analysis, with examples of EMA studies.

Reports a study evaluating actual compliance with written diaries compared with electronic diaries.

Szalai A. 1966. The multinational comparative time budget research project: a venture in international research cooperation. *Am. Behav. Sci.* 10:30

Taylor BC, Fried L, Kenardy J. 1990. The use of a real-time diary for data acquisition and processing. *Behav. Res. Ther.* 28:92–97

Tennen H, Affleck G, Armeli S. 2005. Personality and daily experience revisited. *J. Personal.* 73:1–19

Tennen H, Affleck G, Coyne JC, Larsen RJ, DeLongis A. 2006. Paper and plastic in daily diary research: comment on Green, Rafaeli, Bolger, Shrout, and Reis. *Psychol. Methods* 11:112–18

Thiele C, Laireiter A-R, Baumann U. 2002. Diaries in clinical psychology and psychotherapy: a selective review. *Clin. Psychol. Psychother.* 9:1–37

Todd M, Armeli S, Tennen H, Carney MA, Ball SA, et al. 2005. Drinking to cope: a comparison of questionnaire and electronic diary reports. *J. Stud. Alcohol* 66:121–29

Tourangeau R. 2000. Remembering what happened: memory errors and survey reports. See Stone et al. 2000, pp. 29–48

Tulen J, Volkers A, Stronks D, Cavelaars M, Groeneveld W. 2001. Accelerometry in clinical psychophysiology. In *Ambulatory Assessment*, ed. J Fahrenberg, M Myrtek, pp. 207–31. Seattle, WA: Hogrefe & Huber

Turner JR, Ward MM, Gellman MD, Johnston DW. 1994. The relationship between laboratory and ambulatory cardiovascular activity: current evidence and future directions. *Ann. Behav. Med.* 16:12–23

Tversky A, Kahneman D. 1973. Availability: a heuristic for judging frequency and probability. *Cogn. Psychol.* 5:207–32

Van Den Brink M, Bandell-Hoekstra E, Abu-Sadd H. 2001. The occurrence of recall bias in pediatric headache: a comparison of questionnaire and diary data. *Headache* 41:11–20

Verbrugge LM. 1980. Health diaries. *Med. Care* 18:73–95

Walls TA, Schafer JL, eds. 2006. *Models for Intensive Longitudinal Data*. New York: Oxford Univ. Press.

Wheeler L, Reis H. 1991. Self-recording of everyday life events: origins, types, and uses. *J. Personal.* 59:339–54

Wilhelm FH, Roth WT, Sackner MA. 2003. The LifeShirt: an advanced system for ambulatory measurement of respiratory and cardiac function. *Behav. Modif.* 27:671–91

Willis G. 2005. *Cognitive Interviewing*. Thousand Oaks, CA: Sage

Describes a variety of methods for statistical analysis of EMA data.

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Clinical Psychology* chapters (if any)
may be found at http://clinpsy.AnnualReviews.org

ANNUAL REVIEWS

It's about time. Your time. It's time well spent.

New From Annual Reviews:

Annual Review of Organizational Psychology and Organizational Behavior

Volume 1 • March 2014 • Online & In Print • http://orgpsych.annualreviews.org

Editor: **Frederick P. Morgeson,** *The Eli Broad College of Business, Michigan State University*

The *Annual Review of Organizational Psychology and Organizational Behavior* is devoted to publishing reviews of the industrial and organizational psychology, human resource management, and organizational behavior literature. Topics for review include motivation, selection, teams, training and development, leadership, job performance, strategic HR, cross-cultural issues, work attitudes, entrepreneurship, affect and emotion, organizational change and development, gender and diversity, statistics and research methodologies, and other emerging topics.

Complimentary online access to the first volume will be available until March 2015.

Access this and all other Annual Reviews journals via your institution at www.annualreviews.org.

ANNUAL REVIEWS
It's about time. Your time. It's time well spent.

New From Annual Reviews:

Annual Review of Statistics and Its Application

Volume 1 • Online January 2014 • http://statistics.annualreviews.org

Editor: **Stephen E. Fienberg,** *Carnegie Mellon University*
Associate Editors: **Nancy Reid,** *University of Toronto*
Stephen M. Stigler, *University of Chicago*

The *Annual Review of Statistics and Its Application* aims to inform statisticians and quantitative methodologists, as well as all scientists and users of statistics about major methodological advances and the computational tools that allow for their implementation. It will include developments in the field of statistics, including theoretical statistical underpinnings of new methodology, as well as developments in specific application domains such as biostatistics and bioinformatics, economics, machine learning, psychology, sociology, and aspects of the physical sciences.

Complimentary online access to the first volume will be available until January 2015.

Access this and all other Annual Reviews journals via your institution at **www.annualreviews.org**.