Getting Started: Launching a Study in Daily Life

Tamlin S. Conner — University of Otago

Barbara J. Lehman — Western Washington University

Author Affiliations:

Tamlin S. Conner, Ph.D.
Department of Psychology
University of Otago
Dunedin, 9054
New Zealand
tconner@psy.otago.ac.nz

Barbara J. Lehman, Ph.D.
Department of Psychology
Western Washington University
516 High Street MS 9089
Bellingham WA 98225-9172
USA
barbara.lehman@wwu.edu

Conner, T. S., & Lehman, B. (2012). Getting started: Launching a study in daily life. In M. R. Mehl and T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (89 – 107). New York, New York: Guilford Press.

Getting Started: Launching a Study in Daily Life

Capturing naturalistic human experience is a fundamental challenge for researchers in psychology and related fields. Fortunately, as demonstrated by the broad scope of this book, there are an increasing number of methodologies for studying the full range of experiences, behavior, and physiology as people go about their daily lives. The goal of this chapter is to provide a starting point for researchers interesting in using these methods. Our aim is to provide practical guidance and basic considerations in how to design and conduct a study of individuals over time in their naturalistic environments. We explain how basic design considerations depend on a number of factors—the type of research question, the characteristics of the sample of interest, the nature of the phenomena under investigation, and the resources available to conduct the research. We address each consideration in turn, and integrate our practical discussions with examples that draw on different types of research questions. In this way, our chapter lays a foundation for the more advanced chapters that follow in this section of the Handbook.

Designing a Study to Capture Daily Life: Preliminary Considerations

Figure 5.1 outlines the important steps in designing a study of everyday life. The first steps are to determine the research question, the target variables (experience, behavior and/or physiology), and the population of interest (university students, children, married couples, cancer patients, etc). These decisions, in turn, influence the type of method, sampling strategy, technology platform, and conduct of the study.

Insert Figure 5.1 about here

The Research Question and Target Variables

Every study starts with a well defined research question that determines whether intensive longitudinal methods are appropriate. These methods are geared towards investigating *micro-level processes* at the daily level—that is, the content of and patterns surrounding experiences, behaviors, or physiology as they unfold in real-time or close to real-time in daily life. Accordingly, these methods are best suited to questions aimed at the micro-level such as 'how does mood vary across the course of a week?', 'which individuals cope more effectively with daily stressors?' or 'how do cardiovascular responses differ between the workplace and home environments?' Questions about macro-level processes, such as the relation between lifetime events and lifetime outcomes (e.g., childhood abuse leading to adult depression), are better suited to traditional longitudinal methods in which assessments are taken across months, years, or decades rather than moments or days. Of course, macro-level variables can be examined as predictors of daily micro-level processes (e.g., 'do child abuse victims react differently to daily stressors than non-victims?')—but ultimately, the research question must contain a focus on micro-level daily processes as predictor or outcome.

Intensive longitudinal methods are also better suited to answering descriptive or correlational questions, rather than causal questions. Experimental control is lacking in naturalistic environments; there is no random assignment to different daily circumstances; and everyday events are not manipulated, at least not by experimenters. As a result, these methods are far better suited for understanding the social or emotional circumstances in which a phenomenon occurs, or, the conditional effects of environmental events on psychological processes. Terms like 'predict' and 'associated' are favored over terms like 'cause' and 'effect.' Some elements of causality can be tested using lagged analyses in which events at an objectively earlier time period (e.g., stress during the day) are used to predict later experiences (e.g., drinking that night); however, lagged analyses only indicate precedence, a necessary but not sufficient condition of causality.

The decision to use micro-longitudinal methods also requires acknowledging another caveat— dispelling the myth that momentary assessments of experience and behavior are the 'gold standard' of subjective self-report methods. Historically, as pager-based experience sampling methods were introduced and later invigorated with mobile computers, real-time data capture methods developed a reputation as the 'gold standard' of self-report procedures. The perspective that these methods are inherently better than other types of self-reports (retrospective, global/trait) is outdated. As noted in the chapters by Reis and Schwarz (this volume), experiential reports and retrospective reports provide different types of information— one is not necessarily better than the other. A growing body of research suggests that how people filter and remember their experiences, not necessarily what actually happened across various moments can be the stronger predictor of some future behavior (e.g., Wirtz, Kruger, Scollon, & Diener, 2003). The key determinant is whether the research question concerns the 'experiencing self' or the 'remembering self' (Kahneman, & Riis, 2005) (see also Conner & Barrett, in preparation).

Clarifying the research question, in turn, drives decisions about what variables need to be measured— experiences, behaviors, or physiology; or a combination of these. Table 5.1 shows a proposed taxonomy of the methods best suited to capturing these variables. One key distinction is between active versus passive methods of data collection. Active simply means that the participant is involved in providing the measurement either through self-report (as with daily diaries or experience sampling methods) or through some other voluntary action like giving a saliva sample (as with ambulatory neuroendocrinology). Passive means that the data are collected using devices without any direct involvement from the participant except for wearing the device. The choice between active and passive methods is particularly important when measuring behaviors like physical activity. As noted by Bussman and Ebner-Priemer (this volume), the correlation between self-reported physical activity and activity measured passively through actigraphs is moderate at best, suggesting that these methods capture somewhat different although overlapping phenomena (subjective perceptions of exercise versus objective bodily movements). This active-passive distinction is also relevant to measuring experiences like emotions. From a strict phenomenological perspective, momentary emotional experience can only be assessed using real-time self-report methods like experience sampling, daily diaries, and event-contingent sampling; however, emotions can also be observed and coded from verbal behavior captured using ambulatory acoustic monitoring (see Mehl and Robbins, this volume). As Mehl and Robbins explain in their chapter, these two approaches reflect different perspectives: one from the viewpoint of the self; the other, from the viewpoint of the observer. The key to designing a successful study is to choose the method that maps onto the intended construct—whether through the eyes of the actor or inferred through other means.

Insert Table 5.1 about here

Characteristics of Research Participants

Choices regarding design, research protocol, and the platform for data collection are influenced by the characteristics of the participants. Research approaches that are possible with university students may not be feasible or wise in other samples. Important sample considerations include whether or not all potential participants have consistent internet access, sufficient comfort with technology, and strong verbal skills, as well as whether they can be trusted to follow protocols and care for equipment. The cost of equipment, consequences of equipment loss or damage, and available incentives to aid in equipment care and return should also inform design choices. For example, an internet-based approach that makes use of auditory, pictorial, or visual stimuli might be particularly appropriate for children who have consistent internet access during the times of interest. Adolescents or younger adults might find it most convenient and rewarding to use their own mobile phones for text messaging (e.g., Conner & Reid, 2011). Those who could be trusted with equipment could also use handheld devices like a specialized mobile phone or personal digital assistant (PDA). Depending on the topic, such a study might effectively frame responses like Twitter "tweets" or Facebook status updates. However, these same approaches would not be as appealing or appropriate for a study with elderly adults who might find it difficult to read fine print on a PDA, who may lack internet access, or who may not be accustomed to technology. The human factors involved with using PDAs can be especially problematic for some older individuals. The screens are small and the audible prompts are often at high frequencies, which can make them difficult to hear. For an older adult sample, interactive voice response (IVR) methods through the telephone (see Mundt, Perrine, Searles, & Walter, 1995) or pencil and paper approaches, possibly combined with handheld automated time-date stamps to confirm compliance, may be a more appropriate choice. As younger generations age, access to and comfort with technology will be less of an issue.

Concern for equipment loss would be particularly pronounced if expensive equipment such as PDAs, cell phones, or digital cameras were being used by participants with limited incentive to return equipment. However, Freedman, Lester, McNamara, Milby, and Schumacher (2006) described success in providing inexpensive cell phones to homeless individuals undergoing outpatient treatment for cocaine addiction. Participants reported on cravings and substance use through an IVR initiated each time the participants responded to a randomly initiated cell phone call. Participants were paid in incremental amounts, totaling up to US$188. Only 1 of 30 participants failed to return the cell phone, and participants generally cared for the equipment and followed research protocols.

Likewise, although participant burden should always be minimized as much as possible, a simple, relatively unobtrusive approach would be especially important for highly stressed participants who have limited time and resources. For example, a one week daily diary study is relatively low in terms of participant burden. For stressed but highly responsible participants, passive acoustic monitoring (see Mehl and Robbins, this volume) may be especially effective.

Resources

Budget and the desired speed of data collection also influence the data collection platform. Platforms vary considerably in cost, with the most expensive options drawing on ambulatory monitors of cardiovascular or respiratory functioning (e.g., the LifeShirt) (see Wilhelm and Grossman, this volume) and the least expensive utilizing pencil and paper responses. The Appendix gives an estimate of the relative costs of different approaches. It is best to consider projected costs at the start of a study; this will allow you to choose the platform that accommodates your budget.

Equipment costs can be the most expensive part of these studies, but they do not have to be. For many studies in everyday life, particularly for studies requiring the highest level of control and flexibility, researchers must purchase smart phones, PDAs, tablet computers, or other devices before the study is underway (see Kubiak and Krog, this volume, for device options). Although utilizing a simple once-daily paper and pencil diary instead of a computerized approach costs less, this choice considerably reduces researcher control and is only feasible for some types of research questions. Good low-cost computerized alternatives include using a daily survey that people access through the internet, substituting a person's own cell phone as a signalling device instead of pager, or using a person's own cell phone to conduct experience sampling through text-messaging (e.g., Kuntsche & Robert, 2009) or interactive voice responding (IVR) (e.g., Courvoisier, Eid, Lischetzke & Schreiber, 2010). There are also low cost ambulatory devices like pedometers for measuring physical activity. If greater control is needed or if more expensive equipment is required, it is possible to minimize costs by collecting data on a smaller number of participants at one time over a longer period of time (e.g., purchasing 10 devices and collecting data in waves). It is also possible for researchers to share resources and equipment, and therefore embed data for several testable hypotheses within a single study. Lastly, although it adds to the overall costs, it is useful to purchase some replacement equipment at the start of a study. For example, purchasing an extra 20% (e.g., two extra devices for a set of 10) extends the lifetime of the fleet because units can be replaced individually as they become lost or damaged.

Of course, costs arise not only from equipment expenses but also from payments to research assistants and research participants. Because research in everyday life often involves considerable contact with participants and produces large quantities of data, the research is often conducted by a relatively large team of well-trained researchers. Likewise, participation often requires considerable time and attention to detail. It is therefore useful to compensate participants well and offer appropriate incentives. Studies vary in how they compensate participants. Sometimes researchers tie payment to response rates (e.g., paying for every report made), which is a good strategy for daily diaries, but may not work for more intensive sampling involving multiple reports per day. In these cases, researchers can set a minimum criterion to be met for payment (e.g., completing a minimum of 50% to 75% of all reports), or offer payment or raffle entry based on the percentage of reports made. Remuneration can be supplied in forms other than cash payment, such as University course credit, raffles, gift certificates, and other incentives. Remuneration is a somewhat controversial issue. Researchers should consult with their own ethics boards to determine what remuneration formats provide the appropriate balance of compensation and incentive, without being overly coercive. Offering appropriate incentives throughout the procedure, although not so much that it would be unnecessarily coercive, can proactively reduce attrition and help to protect equipment.

Sampling Strategy and Technology Platform

An essential step in designing a study is to choose the frequency and timing of observations (also called the sampling strategy). The Appendix presents four main types sampling strategies, a summary of when to use them, and the technology platforms available for each approach at the time of publication. As shown in column 2 of the Appendix, sampling strategies depend on the research question and the characteristics of the phenomenon of interest—the type of experience, behavior, or physiology that one wishes to sample. Expected frequency of the phenomenon is the key to deciding the type of sampling strategy. If the occurrence of the phenomenon is relatively rare, it is unlikely that randomly sampled moments throughout the day would effectively capture the behavior. A better strategy for such an occurrence would be for participants to record decisions after the event happens (using event-based sampling). This may require a longer period of time (weeks to months) to allow for sufficient recording of events. In contrast, if the behavior of interest is ongoing (e.g., mood or physiology), then experience sampling throughout the day over a shorter time span (days to weeks) or continuous sampling will work better. It is important for researchers and reviewers to note that most studies of everyday life are designed to randomly sample observations with the intent of generalizing to a *population of occasions*. Although it was controversial at the time, Brunswik (1955) similarly argued that to effectively capture psychological phenomena, researchers must necessarily sample observations from environmental contexts, just as we sample from populations of participants. It is therefore important to sample enough occasions or time periods so that the selected observations can generalize to the population of experiences.

As shown in the Appendix, the first two sampling strategies are ‘time-based’ protocols (Bolger, Davis, & Rafaeli, 2003), where the timing of assessments occur at either variable or fixed times. For variable sampling, also called ‘signal-contingent sampling’ (Wheeler & Reis, 1991), assessments are made in response to a signal that is delivered at unpredictable times, typically between four to 10 times each day. The signals are usually distributed throughout the day “randomly within equal intervals.” For example, for a study sampling eight times a day between 9 AM and 9 PM, the first signal would come randomly between 9 and 10:30 AM, the second signal would come randomly between 10:30 AM and 12 PM, and so on, because there are eight 90 minute intervals in those 12 hours. A minimum necessary time between signals (e.g., 30 minutes) can also be specified so that signals are not too close together. By sampling across the day, this approach aims to generalize across a population of occasions during wakeful hours. At each signal, people may directly report their current experiences (e.g., mood, pain), or, passively have an indirect assessment taken through ambulatory techniques (e.g., blood pressure reading). With variable sampling, the risk for participant burden is high. As such, observations should be frequent enough during each day to capture important fluctuations in experience, but not so frequent so to inconvenience participants without any incremental gain (Reis & Gable, 2000). There are no set rules for the number of sampled moments taken per day, but general guidelines have emerged. The usual range is four to 10 signals per day, with about six being normative. For example, Delespaul (1992) advises against sampling more than six times per day over longer sampling periods (i.e., three weeks or more) unless the reports are especially short (i.e., 2 minutes or less) and additional incentives are provided. Because assessments are frequent and unpredictable, variable schedules are well suited for studying target behaviors that are ongoing and therefore are likely to be occurring at a given signal (e.g., mood, pain, physiology, stress levels). They are also appropriate for studying subjective states that are susceptible to retrospective memory bias (e.g., emotions, pain, or any experience quick to decay), as well as states that people may attempt to regulate if reports were predictable. The main disadvantage of variable time-based sampling is the heightened burden to participants, who are interrupted by the signal. Fortunately participants typically become accustomed to the procedure rather quickly (See *Practical Concerns* below).

For a *fixed timing schedule*, also called interval–contingent sampling (Wheeler & Reis, 1991), assessments are made at set times throughout the day (e.g., at morning, afternoon and evening intervals; or at night daily). Participants may be asked about their experiences or behaviors at that moment (momentary report) or about their experiences or behaviors during the time frame since the previous report (interval report). This latter format requires some retrieval or reconstruction over a period of time and is best used for studying concrete events and behaviors that are less susceptible to memory bias (e.g., a checklist of daily events; reports of exercise, food intake, etc.). One of the most common types of fixed sampling protocols is the daily diary in which experiences and behaviors are reported the end of the day for a period of time (typically one to four weeks). This approach is covered in greater detail by Gunthert and Wenze, this volume. Fixed schedules are also well-suited to time series investigations of temporal phenomena. For example, Courvoisier, Eid, Lischetzke, and Schreiber (2010) utilized a fixed timing schedule to assess six mood measures each day for one week, and evaluated within and between-day fluctuations in mood and compliance with a mobile phone survey. The fixed nature of observations allows one to make generalizations about time (e.g., diurnal or weekly patterns in mood) by statistically comparing responses within- and between-individuals. The timing of the assessments must map onto the nature of the phenomena however. For cyclical temporal phenomena, such as diurnal variations in mood or cortisol responses, the observations need to be frequent enough to catch the trough and peak of the cycle. Otherwise, cycles will be missed or misidentified (an error known as *aliasing* in the time-series literature). Fixed schedules are typically the least burdensome to participants. Reports are made at standardized times and participants can configure their schedules around these reports. This regularity can be a drawback however. If people make their own reports or initiate them in response to a signal at a fixed time, their reports will not reflect spontaneous contents in the stream of consciousness. Reports can also be susceptible to mental preparation and/or self-presentation. If these issues are a concern, then a variable schedule can be used, or, the fixed schedule ‘relaxed’ so that prompts are delivered less predictably around specified times.

A third type of sampling protocol is event-based sampling, in which assessment are made following a pre-defined event. This type of protocol, which is also called event-contingent sampling (Wheeler & Reis, 1991) and event-contingent recording (Moskowitz and Sadikaj, this volume), is best used for investigating experiences and behaviors surrounding specific events, especially those that are rare and may not be occurring if one is sampled at fixed or intermittent times during the day (e.g., instances of conflict, lying, smoking, social events). Event contingent sampling is typically initiated by the person soon after an event has occurred. In these applications, event-contingent sampling can be challenging to participants especially if the events are very frequent or are too broadly defined; so it is important to set clear and appropriately inclusive criteria for the target event. The unique challenges and advantages of event sampling are covered by Moskowitz and Sadikaj, this volume. Recent technological advancements also make it possible to trigger event-recording through other events such as changes in environment (see Intille, this volume) or changes in physiological states such as heightened arousal. Although this technology is still in development, Uen et al. (2003) described the advantages of using ambulatory electrocardiogram readings to trigger an ambulatory blood pressure reading to identify the occurrence of asymptomatic heart irregularities. Likewise it is possible to use ambulatory cardiovascular readings to generate a signal on a PDA or similar device to probe social activities at the time of a physiological trigger (see Wilhelm, Grossman, & Mueller, this volume).

A fourth type of sampling protocol is continuous sampling, in which assessment are made continually without any gaps. Continuous sampling methods are currently most frequently used to address physiological or medical research questions, but with technological advances these approaches are starting to be used in other domains. An advantage to continuous sampling strategies is that the data are captured throughout the duration of the study; nothing is missed. For example, when used in combination with reports on the occurrence of stressful events, continuous electrocardiogram monitoring allows the researcher not only to examine cardiovascular reactivity to the event itself, but also to examine cardiovascular recovery following the event. When continuous sampling is used, it is possible to analyze the entire period of participation, or to sample random or event-based moments from the continuous data stream. In addition to ambulatory cardiovascular readings, researchers have continuously sampled physiological readings such as respiratory function (Houtveen, Hamaker, & van Doornen, 2009), skin conductivity (Thurston, Blumenthal, Babyak, & Sherwood, 2005), and ambulatory glucose levels (Kubiak, Hermanns, Schreckling, Kulzer, & Haak, 2004). The sampling duration of these approaches is often short (1-3 days) and limited by the device's battery life and storage space.

It is also possible to use more than one sampling strategy within a single study, or to use different sampling approaches to address the same general topic. For example, recent research has begun to draw on naturalistic approaches to examine the psychological effects of stigma management. Beals, Peplau, and Gable (2009) used a combination of an event contingent and a fixed time-based approach to study the emotional consequences of opportunities for gay men and lesbians to disclose their sexual orientation. Participants in the two week study completed a report each time they experienced a situation when they felt they could have disclosed their sexual orientation (event-contingent record), and also completed a nightly time-based report describing their social support, coping, and affect over the day (daily diary record). In a study examining a similar topic, Hatzenbuehler, Nolen-Hoeksema, and Dovidio (2009) used a daily diary approach that required participants to provide a nightly report in which they described both stigma-related and other stressors that had occurred during the day. Both African-American and gay male and female participants reported greater emotional distress, more rumination, and more emotional suppression on days when they described stigma-related stressors. These two studies provide examples of how several different sampling strategies may be used to effectively study everyday experiences with stigma-related stress.

Duration of Sampling Period and Statistical Power

The duration of the sampling period is also a key factor in studies of daily life. Studies involving multiple reports per day (variable or fixed) typically run from three days to three weeks. Daily diary studies typically run from one to four weeks. Event-based studies depend on the frequency of the phenomenon with the duration ranging from one week for frequent events (e.g., interpersonal interactions) to several months for less frequent events (e.g., risky sex). Continuous studies are typically conducted over short time periods (i.e., 1-3 days). Continuous studies lasting any longer than a week is currently technically difficult and overly burdensome to participants.

Statistical power plays an important role in deciding the duration of the study and the required sample size. Fundamentally, the number of moments sampled must provide a reliable estimate of the phenomena for each person. We refer the reader to Bolger, Laurenceau and Stadler, this volume, for a more detailed description of statistical power and how it affects the number of observations required per person and sample size. As with most analyses of statistical power, the total number of observations needed per person and the sample size recommendations vary considerably with the complexity of the research hypotheses. For example, researchers who are interested primarily in associations among phenomena being sampled from daily experiences—and not in the role of individual difference factors in moderating those associations—may require relatively fewer research participants.

Technology Platform

The Appendix also lists the most common technology platforms currently available for conducting studies in daily life. The choice of technology platform reflects a trade-off between cost, complexity, and control. Computerized methods cost more and are more challenging to implement, but they provide the greatest control over the timing elements, i.e., by controlling when reports are made and time-date-stamping each report. Platforms for computerized experience sampling include personal digital assistants (PDAs) or smart phones (mobile phones outfitted with specialized configurable software that enables participants to answer questions about their experiences). For a discussion of device and software options see Kubiak and Krog, this volume. Computerized devices can be used with any of the sampling strategies, but they are especially beneficial for time-based protocols because compliance with timing can be validated. However, the use of automatically recorded times does not ensure compliance with event-based procedures in which participants initiate their own reporting. Although a self initiated report is time-date stamped, researchers cannot know objectively how much time has passed since the event. Delays between the occurrence of the event and the reporting on it have the potential to introduce memory bias. Computerized devices have several other advantages over non-computerized approaches including the capacity to randomize how questions are presented (to reduce response sets and order effects) and the capacity to record useful ancillary information such latencies to respond to each question. The disadvantages start with the upfront financial investment. The cost of units can vary considerably. There is also the potential cost for the software (which is free to very costly; see Kubiak and Krog, this volume). Computerized platforms also often impose limits on the question format. Open ended or “free” responses are not easily incorporated except perhaps for longer internet based daily diaries. Finally, a computerized platform may not be suitable for all subject populations (see *Characteristics of Research Participants* section above).

There are several lower cost alternatives to computerized sampling with a fleet of PDAs or smartphones. As previously mentioned, experience sampling using participants own mobile phones through texting or interactive voice response (IVR) systems is becoming increasingly popular. With texting, researchers send texts to participants' personal mobile phones using an in-house server or commercial SMS (short messaging service) company for a small cost per text (Conner & Reid, 2011; Kuntsche & Robert, 2009). Participants receive the texts and reply to a set number of questions (either presented in the text or in a separate key) using their numeric keypad. This approach is relatively affordable and can yield high compliance rates. For example, in a two-week texting study of New Zealand university students (using a local SMS company, www.message-media.com.au), participants replied to on average 95% of six texts per day (*SD* = 11%; *Range* = 9% to 100%) (Conner & Reid, 2011). Only four participants failed to meet minimum texting criteria (set fairly high at 75% of texts). Removing those cases from analyses, the final compliance rate was quite good (*M* = 96%, *SD* = 5%, *Range* = 77% - 100%). Another option is to use an IVR system in which automated specialized software like SmartQ from Telesage is used to call participants on their mobile phones to complete a survey (see Courvoisier et al., 2010). Of the two options, texting has somewhat less control over timing elements. Although a text may be sent to participants at a certain time, they may not receive it or reply to it immediately (if their phone is turned off or they have poor reception). This may not be too much of an issue as Conner and Reid (2010) found that the median time to reply to texts was only two minutes (*M* = 16 min; *SD* = 37 min; *Range* = <1 min to 9 hours). Both texting and IVR approaches also typically have less space for questions than PDAs or smartphones. If space is an issue, it is possible to use participants' own mobile phones as a low cost pager to remind them to complete a more extensive paper-and-pencil survey. Lastly, Internet based surveys are another low cost alternative, and are especially well suited to a daily diary procedure. Surveys can be designed using relatively low cost software (e.g., www.surveymonkey.com). Every day participants can log on to a secured web site and complete their report, which is time-date stamped and stored for analysis. See Gunthert and Wenze for more detail on internet daily diaries.

Paper-and-pencil surveys are still a low-cost alternative to computerized approaches. In this approach, people carry around booklets or complete a survey each evening prior to going to bed. The advantages of paper-and-pencil methods include reduced cost, less overhead in terms of equipment, and the allowance of open-ended questions. The main disadvantage is the inability to objectively confirm that surveys were completed at the designated times. Although there is evidence that people may not complete paper-and-pencil reports according to the proper schedule (Stone, Shiffman, Schwartz, Broderick, & Hufford, 2002), there is also evidence that paper-and-pencil methods are valid and can be equally informative (Green, Rafaeli, Bolger, Shrout, & Reis, 2006). This “Paper or Plastic” debate has clarified several things. First, it is now acknowledged that computerized methods are better for circumstances necessitating precise timing control and assurance; however, extreme concerns with paper-and-pencil questionnaires may be overstated. Nonetheless, we expect that researchers in the future may find it increasingly difficult to publish research papers that use paper-and-pencil questionnaires if a suitable computerized method is available (and appropriate for that population).

Several different approaches can help promote compliance with paper-and-pencil questionnaires. A simple and often effective approach is to establish a good working relationship with participants and to collect surveys frequently (see Green et al., 2006). It is also possible for even the most low-tech of data collection platforms to include an electronic method for assuring compliance. For example, in their study of everyday stress among adolescents, Fuligni and colleagues (2009) asked participants to complete nightly checklists each day for two weeks. The researchers assured compliance by providing participants with an electronic stamp (DateMark by Dymo Corporation, Stamford, CT) to place across the seal of an envelope holding that day's checklist. This approach is far more convenient, and would likely yield greater compliance than requiring participants to place daily surveys in the mail. However, until the costs of electronic stamps are reduced (cost was US$75 at publication), this might not be cost effective. Another similar option uses time and date stamps recorded by the Medication Event Monitoring System (MEMS Smartcap; Aardex Group, Union City, CA), a small container that can hold pills or supplies necessary for research participation (e.g., materials to provide saliva samples). For example, in a two-week study, Applebaum and colleagues (2009) examined discrepancies between MEMS pill bottle openings and HIV-infected participants' self-reported adherence with antiretroviral medications; participants reported greater medication adherence than was documented by the MEMS caps.

Designing a Study to Capture Daily Life: Practical Concerns

Participant Issues

Once participants are recruited, it is important to keep them motivated through the study. There are several strategies for maintaining motivation including having a complex remuneration structure with incentives such as money, research credits, and lotteries. Positive attitudes from research assistants are also crucial. Through the authors' and others experiences (see Green et al., 2006), data quality appears to be highest when participants feel they are valued, have a sense of responsibility to the research assistant, and believe that the research itself is important. Participants should also clearly understand the study procedures. They should know how and what they will be asked to report, when (roughly) they will report, and how to use the computerized device. With devices, it is best to have participants answer their first prompt in the lab, giving them the opportunity to ask questions and get comfortable with the technology. Also, during the study, participants can be given feedback on their response rate to improve compliance. For especially complex studies, it may be useful to have a cell phone “hotline” that participants can call if they have questions or equipment difficulties.

Data Preparation and Analysis

Each type of method and platform presents its own unique challenges to data preparation. More detailed guidelines for data cleaning are covered by McCabe, Mack, and Fleeson, this volume. With paper-and-pencil methods, all data must be entered manually by hand—a lengthy and error-prone process. Computerized approaches bypass this step because data are retrieved directly from the portable device, internet, or the text-messaging server; however, they still need to be cleaned and organized. For example, with text-messaging based studies, records need to be checked for duplicates that occur when participants inadvertently send their text responses twice. Likewise, if time and date stamps are used to match participant self-reports with their physiological responses (e.g., cortisol samples or blood pressure readings), these cases must be matched up prior to analyses. Researchers should take care to assure that equipment always has the correct time and date. It is also important to be aware of special date-time transitions, such as daylight savings time and leap years. Moreover, in device-driven sampling, if reaction times are measured, trials with extremely fast reaction times typically indicate participant error such as an inadvertent screen tap and should be removed (See McCabe et al., this volume for recommendations). Unusually fast responding (<500 milliseconds, see McCabe et al., this volume) can also sometimes indicate inauthentic responding, particularly if the person gives the same response for each item. Individuals with no variability in their responses should also be flagged and examined. Although there are some measures where no variability would be expected (e.g., a food diary in which a person reports eating no chocolate every day), for other phenomena like mood, stress, and other varying experience, no variability may indicate a response set. These cases can be detected by computing a within-person standard deviation for each item across all days for each individual. Any participants removed for non-compliant responding should be noted in the write up.

After data have been checked and prepared, they are ready to be analyzed. The chapters in Section III of this Handbook cover the common ways of analyzing intensive longitudinal data. Before beginning analyses, decisions need to be made regarding treatment of missing data. We recommend reading the Handling Missing Data chapter by Black, Harel, and Matthews (this volume). It is important to know why data may be missing and whether the patterns of missing data are random or reflect systematic bias. It is also common practice to delete individuals from analysis if they have excessive missing data. Although minimum response rates are somewhat arbitrary and depend on the amount of data needed for analysis, requiring participants to complete at least 50% of reports seems commonplace. However, as noted by Black and colleagues, such casewide deletion may be too stringent of an approach.

At their core, all analytic methods in Section III recognize the ‘nested’ nature of intensive longitudinal data, whereby observations are nested within people. This nested data structure requires proper analytic treatment to take into account shared variance between observations from the same individual. Multilevel modeling is one of the most common approaches to analyzing nested data (see Nezlek, this volume). It allows researchers to model certain patterns within each individual and to test whether those patterns differ as a function of person characteristics. It should be noted that observations taken closer together in time (10 am and 12 pm on Monday) are typically more similar to each other than they are to other readings from a particular person (10 am Monday vs. 7 pm Friday). This phenomenon of serial autocorrelation can be statistically modelled through a number of approaches, including entering the preceding observation as a covariate in the multilevel model. For examples of studies that modelled serial autocorrelation, see Courvoisier, Eid, Lischetzke, and Schreiber (2010) or Lehman and Conley (2010). Dyadic data analysis takes multilevel modelling one step further and models shared variance between couples and other dyads (see Laurenceau and Bolger, this volume). Other methods allow for more complex model building and draw on techniques such as structural equation modelling and multilevel meditational analyses (Eid and Courvoisier, and Card, this volume). Analytic approaches can also be used to test for complex patterns of change and temporal dynamics (see chapters by Ebner-Priemer, and by Deboeck, this volume) as well as the structure of daily experiences through within-person factor analyses (See Ram, this volume).

Reporting Guidelines

When publishing daily life research, researchers are encouraged to include certain types of information in their write up. Here, we highlight the most important guidelines from Stone and Shiffman (2002). (1) Explain and justify the sampling strategy. Describe the type of sampling strategy; the frequency, timing, and length of the study; and briefly justify these design decisions. (2) Explain the details of the data collection platform. Give a physical description of any computerized devices and software, including the website for software if applicable. Explain how items were presented through the device, whether items were randomized, and whether there was any time limit for responding. (3) Describe any participant training, monitoring, and compensation. Did they receive any training for the devices; did they receive feedback on response rates or have other incentives for enhancing compliance? What other steps were taken to ensure compliance? How were they compensated? (4) Provide detailed compliance statistics.

State the rationale for including or excluding people from analyses; indicate how many people were excluded and why. Report descriptive statistics for the response rates (*M*, *SD*, *Min*, *Max*) either in numbers or percentages, as in “Participants responded to on average 65 out of 84 prompts, reflecting a response rate of 77% (*SD* = 10%, *Range* = 50% to 98%).” Be clear whether these statistics included or excluded people eliminated for low or non-compliant responding. Timing statistics may also be appropriate for some studies. For example, text messaging studies should report descriptive statistics for time delays between when the texts were sent and when the replies were received if such information is available. (5) Discuss any important data management steps taken beyond simple data cleaning procedures (eliminating duplicates, etc). Examples include decisions made resulting in the retention or dropping of some cases (e.g., dropping reports made outside designated time intervals).

Ethical Considerations

In describing an in-depth interdisciplinary study of the daily life of dual career families (c.f., Campos, Graesch, Repetti, Bradbury, & Ochs, 2009), an article in the New York Times (Carey, May 22, 2010) described the study as “oddly voyeuristic,” noting that the study documented “… every hug, every tantrum, every soul-draining search for a missing soccer cleat” experienced by the families who participated in the research. In truth, the approaches described in this Handbook are somewhat unique in their portability and potential intrusion into the personal and psychological space of participants. For this reason they are sometimes met with skepticism from those who are initially learning of the approach. In reality, participants typically habituate relatively rapidly to providing multiple reports or measurements over the course of a day or week, and rapidly become less aware they are being observed. In our experience, participants often report expecting the study to be more intrusive than it in fact was. Research participants also may be motivated by the opportunity for self-reflection and assessment that participation promotes.

However, because studies of everyday life do often require participants to reflect on their own emotional, social, and other experiences throughout several days, several ethical considerations become important. First, it is important to balance the requirements for frequency sampling and study duration against participant intrusiveness and burden. The study should include enough assessments to effectively capture the phenomena of interest, but should always attend to minimizing overall participant burden. Pilot testing can help to ascertain the frequency of occurrence of the phenomenon of interest, and power analyses can be conducted to help determine the optimal sample size and frequency of assessment (see Bolger, Laurenceau, and Stadler, this volume). In addition, pilot testing should help to assure that the assessments are as brief and clear as possible. It is also a good idea to question participants on the burden of participation and obtain suggestions for improvement after the study ends. Second, as with any study, participants should be reminded of their right to discontinue participation without any negative consequences. Third, it may be necessary to provide letters or other gain supplementary approvals to avoid potential negative consequences of research participation. For example, undergraduate student participants may benefit from letters to teachers or work supervisors that explain that participation in the experience sampling study will require the student to use a PDA or a cell phone to respond to a brief survey at random times throughout the day. This letter will make it clear that participants are not texting a friend, but rather are participating in a research study.

Participation in an in-depth experience sampling study may lead participants to reflect on their own emotions and social interactions. In fact, brief assessments of emotional states and thought processes are sometimes used as a clinical tool for promoting self reflection and personal growth (Harmon, Nelson, & Hayes, 1980). Because of these factors, a careful, skilled debriefing session must be an important component of these studies. The session might usefully begin simply by acknowledging that it is possible (and likely) that participation in the study promoted an unusual level of self-focus, and may have touched on sensitive concerns or made participants feel as though their lives were under a magnifying glass. It is a good idea to provide community or campus resources available for psychological consultation if the participant feels it would be useful to continue to explore these topics. If physiological assessments were taken, it may also be useful to provide participants with a summary report of their biological readings over the duration of the study. For example, because ambulatory blood pressure is prognostic for future disease, participants might be instructed to give the readings to their doctor for interpretation or as a "healthy baseline."

Reactivity effects should also be considered. These methods are unique in that they require people to actively attend to and verbalize their experiences and behaviors repeatedly over time. This raises the question about whether intensive self-reporting changes the very phenomena being measured ('reactivity'). Studies from the medical literature show little or modest reactivity effects, at least for self-reported pain (as in Stone, Broderick, Schwartz, Shiffman, Litcher-Kelly, & Calvanese, 2003); however, there are circumstances when reactivity is more likely to occur (e.g., when participants are motivated to change; when monitoring only one experience or behaviour; when testing certain individuals, Conner & Reid, 2011). These circumstances and other issues of reactivity are discussed in greater detail by Barta, Tennen, and Litt, this volume.

Concluding Thoughts

Part of the appeal of intensive daily life methods is the ability to use technology to capture human lived experience. A side effect of this approach is that in a short amount of time, many of these technologies become outdated. Therefore, it seems prudent to conclude that an important part of successful research is being aware of technological advances, and thinking creatively about how to use these new technologies best to answer a particular theoretical question in a particular population. The growth and spread of wired and wireless communities and expanding access to Global Positioning System (GPS) tracking will undoubtedly allow for the immediate upload of participant responses through wireless communication. As researchers, we need to stay connected to these advances and be innovative enough to apply them.

References

Applebaum, A.J., Reilly, B.A., Gonzalez, J.S., Richardson, M.A., Leveroni, C.L., & Safren, S.A. (2009). The impact of neuropsychological functioning on adherence to HAART in HIV-infected substance abuse patients. *AIDS Patient Care and STDs, 23*, 455-462.

Beals, K.P., Peplau, L.A., & Gable, S.L. (2009). Stigma management and well-being: The role of perceived social support, emotional processing, and suppression. *Personality and Social Psychological Bulletin, 35*, 867-879.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579-616.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*, 193-217.

Campos, B., Graesch, A. P., Repetti, R., Bradbury, T., & Ochs, E. (2009). Opportunity for interaction? A naturalistic observation study of dual-earner families after work and school. *Journal of Family Psychology, 23*, 798-807.

Carey, B. (May 22, 2010). Families' every fuss, archived and analyzed. The New York Times.

Conner, T. S., & Barrett, L. F. (in preparation). Trends in ambulatory monitoring of self-reports. *Psychosomatic Medicine: Special Issue on Ambulatory Monitoring (*T. Kubiak and A. Stone, Eds.).

Conner, T. S., & Reid, K. (2011). Intensive momentary reporting of happiness. Manuscript under review.

Conner, T. S., Tennen, H., Fleeson, W. & Barrett, L. F. (2009). Experience sampling methods: A modern idiographic approach to personality research. *Social and Personality Psychology Compass, 3*(3), *292-313.*

Courvoisier, D. S., Eid, M., Lischetzke, T, & Schreiber, W.H. (2010). Psychometric properties of a computerized mobile phone method for assessing mood in daily life. *Emotion*, 10, 115-124.

Delespaul, P. A. E. G. (1992). Technical note: Devices and time-sampling procedures. In M.W. de Vries (Ed.) *The Experience of Psychopathology: Investigating Mental Disorders in their Natural Settings* (xvii, pp. 363-373). New York, N.Y.: Cambridge University Press.

Freedman, M.J., Lester, K.M., McNamara, C., Milby, J.B., & Schumacher, J.E. (2006). Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. *Journal of Substance Abuse Treatment, 30,* 105-111.

Fuligni, A.J., Telzer, E.H., Bower, J., Cole, S.W., Kiang, L., & Irwin, M.R. (2009). A preliminary study of daily interpersonal stress and c-reactive protein levels among adolescents from Latin American and European backgrounds. *Psychosomatic Medicine, 71*, 329-333.

Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods, 11,* 87-105.

Harmon, T. M., Nelson, R.O., & Hayes, S. C. (1980). Self-monitoring of mood versus activity by depressed clients. *Journal of Consulting and Clinical Psychology, 48,* 30-38.

Hatzenbuehler, M.L., Nolen-Hoeksema, S., & Dovidio, J. (2009). How does stigma get under the skin: The mediating role of emotion regulation. *Psychological Science, 20*, 1282-1289.

Houtveen, J.H., Hamaker, E.L., & van Doornen, L.J.P. (2009). Using multilevel path analysis in analyzing 24-h ambulatory physiological recordings applied to medically unexplained symptoms. *Psychophysiology, 47*, 570-578.

Kahneman D., & Riis, J. (2005). Living, and thinking about it: Two perspectives on life. In F.A. Huppert, N. Baylis & B. Keverne (Eds.), *The science of well-being* (pp. 285-304). Oxford: Oxford University Press.

Kubiak, T., Hermanns, N., Schreckling, H.J., Kulzer, B., & Haak, T. (2004). Assessment of hypoglycaemia awareness using continuous glucose monitoring. *Diabetic Medicine, 21*, 487-490.

Kuntsche, E., & Robert, B. (2009). Short message service (SMS) technology in alcohol research: A feasibility study. *Alcohol and Alcoholism, 44*, 423-428.

Lehman, B. J., & Conley, K. C. (2010). Momentary reports of social-evaluative threat predict ambulatory blood pressure. *Social Psychological and Personality Science, 1*, 51-56.

Mundt, J. C., Perrine, M. W., Searles, J. S., & Walter, D. (1995). An application of interactive voice response (IVR) technology to longitudinal studies of daily behavior. *Behavior Research Methods, Instruments, & Computers, 27*, 351-357.

Reis, H. T., & Gable, S. L. (2000). Event sampling and other methods for studying daily experience. In H. T. Reis and C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190—222). New York: Cambridge University Press.

Stone, A.A. & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, *24,* 236-43.

Stone, A.A., Broderick, J. E., Schwartz, J. E., Shiffman, S., Litcher-Kelly, L., and Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction, *Pain, 104*, 343-351.

Stone, A.A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2002). Patient non-compliance with paper diaries. *British Medical Journal, 324*(7347), 1193-1194.

Thurston, R.C., Bulumthal, J.A., Babyak, M.A., & Sherwood, A. (2005). Emotional antecedents of hot flashes during daily life. *Psychosomatic Medicine, 67*, 137-146.

Uen, S., Baulmann, J., Düsing, R., Glänzer, K., Vetter, H., & Mengden, T. (2003). ST-segment depression in hypertensive patients is linked to elevations in blood pressure, pulse pressure and double product by 24-h Cardiotens monitoring. *Journal of Hypertension, 21*, 977-983.

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality. Special issue: Personality and Daily Experience, 59*(3), 339-354.

Wirtz, D., Kruger, J., Napa Scollon, C., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science, 14*(5), 520-524.

Table 5.1.

A Taxonomy of Intensive Longitudinal Methods

	Experience	**Behavior**	**Physiology**
	Phenomenology like mood, pain, fatigue, cognitions, perceptions, and appraisals	Actions that are observable to others like drinking, smoking, exercise, talking, eating, interactions, and location	Internal workings of the body and brain like temperature, breathing, heart rate, blood pressure, and hormone levels
Active	Experiences are self-reported by participants e.g., mood ratings e.g., pain ratings e.g., stress ratings *Experience Sampling* *Daily Diaries* *Event Sampling*	Behaviors are self-reported by participants e.g., self-reported alcohol use e.g., food/exercise diary e.g., event-recording *Experience Sampling* *Daily Diaries* *Event Sampling*	Physiology is measured by participants e.g. participant takes saliva samples which are assayed for cortisol by experimenters *Neuroendocrine Sampling* *Physiological Sampling*
Passive	Experience is inferred through observation. e.g., unobtrusive auditory sampling with the Electronically Activated Recorder (EAR) *Acoustic Sampling*	Behaviors are measured with no intervention or reporting necessary e.g., pedometer or actigraph to infer physical activity e.g., unobtrusive auditory sampling with the EAR e.g., GPS to measure location *Activity Sampling* *Acoustic Sampling* *Passive Telemetrics* *Context Sampling*	Physiology is measured with no intervention by participant e.g., passive sampling of heart rate and blood pressure e.g., continuous glucose monitoring e.g., temperature tracking e.g., measurement of breathing *Neuroendocrine Sampling* *Physiological Sampling*

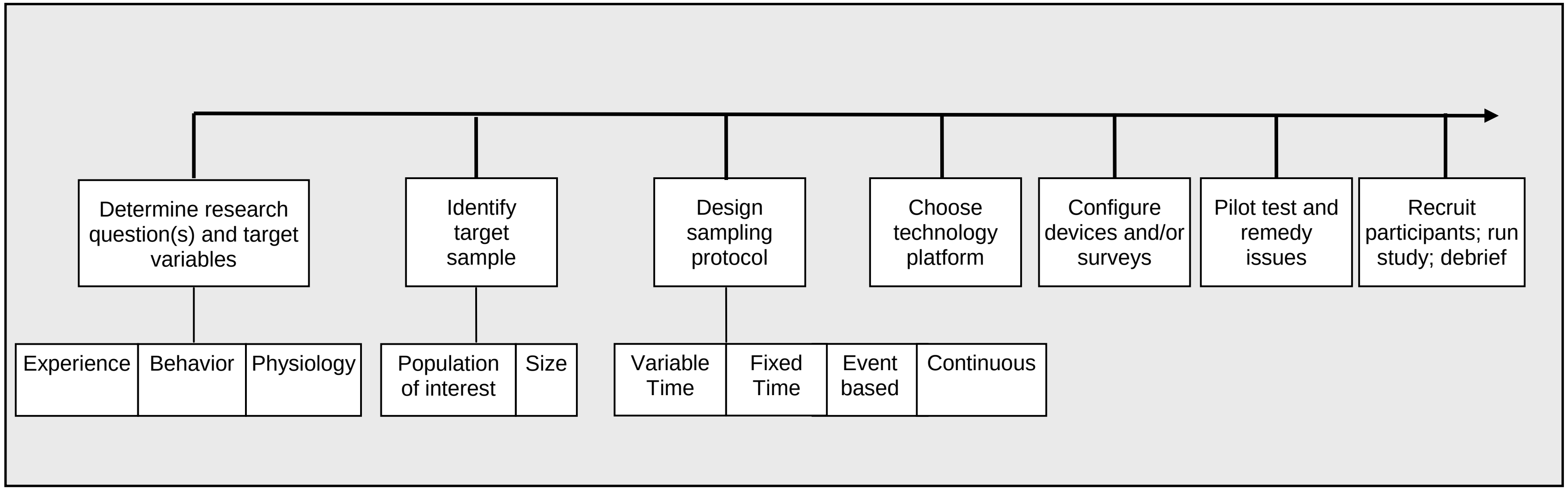

Figure 5.1. Key steps in designing and implementing a study of daily life.

Appendix

Sampling Strategies and Technology Platforms for Data Collection

Sampling Strategy	When to Use	Participant Requirements	Technology Platform for Data Collection	Cost	Participant Responsibilities	Researcher Control[1]
Variable Time-Based						
Reports are made in response to a semi-random signal during the day (4 - 10 times daily); signal times are unknown. Known as *experience sampling* when measuring self-reported experience.	For momentary experiences that are ongoing (like mood), susceptible to memory bias, or may be adversely affected by knowing when a report will be made.	Ability and motivation to provide report when signalled; may require comfort with technology	Personal Digital Assistant Tablet computer Mobile Phone (software) Mobile Phone (texting/IVR[2]) Paper booklet with signalling device (pager, watch, text message)	$$$ $$$ $$$ $-$$ $	equipment care equipment care equipment care mobile access compliance	* * * * * * * * * * * * * * * * * *
Fixed Time-Based						
Reports are made at fixed times (e.g., 10am; 2pm; 5pm or once nightly); reporting times are known and anticipated. Once-a-day reports known as *daily diary methods*.	For experiences and behaviors that are less susceptible to memory bias, and are able to be recalled over prior interval. Also suitable for time-series investigations (e.g., circadian rhythms or daily routines).	Ability and motivation to provide routine reports; may require comfort with technology	Personal Digital Assistant Tablet computer Mobile Phone (software) Mobile Phone (texting/IVR) Telephone call in to IVR Internet Survey E-mail Paper booklet (maybe with time stamp)	$$$ $$$ $$$ $-$$ $$ $ $ $-$$	equipment care equipment care equipment care mobile access phone access internet access internet access compliance	* *
Event-Based						
Reports are made following a predefined event.	To measure processes surrounding discrete events, particularly events that occur infrequently.	Attention to event, and motivation to report event as defined by researcher	Personal Digital Assistant Tablet computer Mobile Phone (software) Mobile Phone (texting/IVR) Telephone call in to IVR Paper booklet	$$$ $$$ $$$ $-$$ $	equipment care equipment care equipment care mobile access phone access compliance	* * * * * * * * * * * * * *

Continuous						
Activities are recorded continuously over a specified time period. Possible to use events from continuous recording (movement, heart rate, etc.) to prompt another mode of data collection.	To measure physiological symptoms or observable experiences that are ongoing and can be captured via ambulatory technology; used to capture naturally occurring events for later coding and analysis.	Must leave monitoring equipment in place and activated	Voice/audio recording (EAR,[3] or other device)	$$$	equipment care	* * * * *
			Lifeshirt (VivoMetrics)	$$$$$	equipment care	* * * * *
			Mobile video recording	$$$$	equipment care	* * * * *
			GPS location monitoring,	$$$	compliance	* * * * *
			portable actigraph, or	$$	compliance	* * *
			pedometer	$$	compliance	* * *

Notes. [1] Researcher control refers to whether the researcher can control and confirm the exact dates and times a report or observation was made. [2] Interactive Voice Recording (IVR) is a phone based system that presents and records answers to survey questions. [3] The Electronically Activated Recorder (EAR, see Mehl and Robbins, this volume) records acoustic data (voice, ambient sounds, etc) for a certain percentage of the day (e.g., 50 sec every 9 min). While not technically continuous sampling, the observations are frequent enough to warrant grouping it with other continuous sampling methods. Portions of this table appeared in Conner, Tennen, Fleeson, and Barrett (2009).