Capturing Momentary, Self-Report Data: A Proposal for Reporting Guidelines

Arthur A. Stone, Ph.D.
State University of New York at Stony Brook

Saul Shiffman, Ph.D.
University of Pittsburgh and invivodata, inc.

ABSTRACT

Self-report data are ubiquitous in behavioral and medical research. Retrospective assessment strategies are prone to recall bias and distortion. New techniques for assessing immediate experiences in respondents' natural environments (e.g., Ecological Momentary Assessment [1], Experience Sampling [2]) are being used by many researchers to reduce reporting bias. This article discusses seven aspects of momentary research that are often overlooked or minimized in the presentation of momentary research reports, yet that are critical to the success of the research: (a) the rationale for the momentary sampling design, (b) the details of momentary sampling procedures, (c) the data acquisition interface, (d) rates of compliance with the sampling plan, (e) the procedures used to train and monitor participants, (f) data management procedures, and (g) the data analytic approach. Attention to these areas in both the design and reporting of momentary research studies will not only improve momentary research protocols but also allow for the successful replication of research findings by other investigators.

(**Ann Behav Med** 2002, *24*(3):236–243)

INTRODUCTION

Self-reports of one's own behaviors, circumstances, and subjective experiences are commonly used in clinical practice and research (3). Some of the more salient examples of self-reported concepts are quality of life and subjective well-being, both of which have been adopted as general endpoints for the evaluation of a wide variety of important societal concerns. Subjective health-related symptoms such as pain, fatigue, malaise, and energy levels often are employed as endpoints in medical research. Reports of behaviors, ranging from food or drug consumption through sexual activities, have been used as both predictors and outcomes in research. Although self-report is often used to assess subjective states, self-report methodologies also are used to collect data about objective events that are most easily or conveniently observed by respondents. It may be safely concluded that self-report data play a significant role in activities that are central to clinical practice and research. In this report, we provide a brief overview of the rationale for momentary studies and then review issues in the reporting of momentary data and suggest guidelines for reporting.

A BRIEF RATIONALE FOR MOMENTARY STUDIES

Much has been written about the challenge of collecting accurate and detailed self-reports of experiences with our most current assessment methodologies—retrospective questionnaire and interview methods. Of the various difficulties, the most vexing is the extensive memory distortion that pervades retrospective self-reports. Cognitive science has shown that the details of past experiences are neither fully encoded into memory nor fully decoded at the time of recall in unbiased ways (4). Autobiographical memory research has shown numerous ways in which recalled information is distorted and transformed by memory and summarization processes (5). In particular, this research suggests that retrieval and interpretation of past experience is heavily influenced by respondents' *current* state; for example, current pain influences recall of past pain (6,7). More broadly, some experts consider recall more akin to reconstruction than retrieval, and it is thus subject to numerous biases at the time of recall (8). Many research and clinical inquiries require the respondent not only to recall past experience but also to summarize it (e.g., How many times did that happen? How severe was it, on average?). The cognitive processes involved in summarizing can introduce further biases (e.g., giving undue weight to more salient or more recent experiences). These influences result not only in random error but also in systematic bias, particularly when the context of recall is influenced by the research design (e.g., when recall occurs in a particular environment such as a research unit.) Several articles and books are available that discuss these memory processes and their effects on recalled information (3,5,9).

A set of techniques has been developed to acquire self-report information with less distortion than found in traditional recall methodologies. These new methods reduce distortion caused by recall bias by assessing phenomena through instantaneous reports of immediate experience. Momentary assessments—that is, assessments that ask a person to report about how he or she feels or what he or she is doing *at the moment*—minimize recall bias because the cognitive processes responsible for distortion should be reduced. These methods also circumvent the need for summary processing because none is being requested of respondents—just a report of their immediate state. In lieu of summaries made by participants, these methods use aggregations of multiple momentary reports to characterize experience over a given period. For example, instead of asking patients to recall how much pain they experienced over the last

We thank Dr. Joseph Schwartz and Dr. Michael Hufford for comments on an early draft of this article.

Reprint Address: A. A. Stone, Ph.D., Department of Psychiatry, SUNY at Stony Brook, Stony Brook, NY 11794–8790.

week, as is common in research practice and clinical care, momentary techniques might instead average randomly selected momentary reports of pain taken several times a day for the entire week (see 10,11). The reports are collected in people's natural environments to sample real-world experience and to avoid the bias potentially introduced by artificial reporting contexts. The family of techniques that use these strategies is often referred to as the Experience Sampling Method (ESM) (2) or as Ecological Momentary Assessment (EMA) (1). EMA is a more broadly defined construct than ESM. ESM has traditionally focused on random time sampling of experience, one important way of characterizing experience. EMA also includes other approaches, such as the recording of events. Although ESM has traditionally focused on private, subjective experience, EMA also explicitly includes self-reports of one's own behaviors and physiological measures.

EMA methods rely on sampling people's experiences. Although it might be preferable to have continuous online measures of self-reports and to summarize them, no methodology exists for obtaining such reports without extreme burden and intrusiveness for study participants or patients. The emphasis on sampling is an essential element of EMA. Momentary experience is expected to vary over time and to be influenced by a number of factors, some that may be relevant to a given study (e.g., medication regimen) and others that may not be (e.g., weather). Thus, the characterization of a person depends on representative sampling of momentary states to estimate the whole, much in the way that representative sampling of participants is seen as essential for valid inferences from a sample to a population of people.

Momentary data collection methods are increasingly appearing in literature as a method to capture people's experiences in the real world as they go about their everyday activities. The scope of these methods ranges from complex sampling schemes that assess participants several times a day to daily diaries completed once a day that are intended to avoid or reduce the problems of retrospection. To date, several special issues of scientific journals have been devoted to the topic, and several books, many book chapters, and dozens of articles have used the methodology (for overviews, see 3,12).

GUIDELINES FOR REPORTING MOMENTARY STUDIES

Unlike other well-established methodologies, few standards or guidelines are available to counsel those contemplating using momentary methods and to inform the consumer of momentary research to determine whether techniques were properly implemented. As in any other area of methodology, understanding the details of the method is key to interpreting the resulting data; this requires that research reports include the relevant information. Research reports must meet two important criteria that go to the function of publication: First, they must allow the reader to interpret the findings in light of the particular research methods, including making comparisons across studies. Second, they must allow the reader, at least in principle, to replicate the study procedures. Meeting these objectives requires detailed reporting of methods and methodological findings. We propose a set of criteria that might serve as a starting point for informing users and consumers about momentary data collection methods. In doing so, we discuss many technical and conceptual aspects of conducting momentary research to show the importance of considering these topics and ensuring that descriptions of momentary protocols provide a level of detail that enables replication of the protocol. We hope these guidelines enable more effective communication among investigators conducting momentary research and more informed design and evaluation of momentary studies.

Although our primary focus is on standards for *reporting* EMA studies, this is inextricably tied to standards for *performing* EMA studies. In the course of proposing reporting standards, we describe why we believe certain aspects of the design and methods are crucial to the validity and interpretation of EMA research. Taking on this task, we understand that the guidelines presented will undoubtedly reflect our own emphases and concerns about conducting momentary studies. We fully expect readers of this article to form their own opinions about what is most important in the conduct and reporting of momentary research, and we believe that this will help the field advance. To aid readers of this report, we have summarized the major points discussed in Table 1.

Provide the Rationale for the Momentary Sampling Design

One of the first and most basic questions an EMA researcher faces in constructing study design is the plan for sampling moments. It is essential to begin this task with a well-thought-out set of research questions and then to fashion the sampling scheme to match those questions. An example is presented to show how specific hypotheses inform the development of the sampling scheme. More detail about sampling strategies can be found in Delespaul (13) and Reis and Gable (12).

We posit a hypothetical study examining daily level of pain in two groups in which one group receives a new treatment for pain reduction and the other group receives placebo treatment. The key to this study hinges on the idea that "daily level of pain" is the desired outcome measure. An EMA approach aims to assess daily experience by obtaining a representative sample of moments throughout the day. The reason for this approach involves two key factors. The first is based on our preference for a momentary approach to reduce recall bias, which means that we will ask about pain experienced at the moment as opposed to the recall of pain over several hours. Because we expect that pain will vary considerably from moment to moment (because of a host of psychological, biological, and environmental factors), we believe we should sample multiple moments so that our summary measure adequately represents the pain experienced throughout the day. Representatively sampling moments, the second factor, is necessary so that our samples are not entrained to particular situations or conditions that might be systematically related to pain level. For example, daily sampling at noon could bias reports given the high probability that situational factors associated with lunch could influence the reports.

TABLE 1
Summary of Design Issues and Reporting Guidelines

Design Issue	*Reporting Guidelines*
Sampling	a. Rationale for sampling design b. Sampling density and schedule c. Implications for bias and validity
Momentary procedures	a. Description of prompting and recording methods b. Description and definition of participant-initiated event entries c. Description of how nonresponses are handled d. Whether response-delaying procedures were available e. Definition of an *immediate* and *timely* response
Data acquisition interface	a. Description of physical characteristics of diary or palmtop computer b. Description of mode of item presentation c. Discussion of important algorithm features d. Text of items and response options and how they were derived or modified
Compliance	a. Rationale for compliance decisions b. Presentation of systematic compliance rates (% of required assessment episodes completed) c. Demonstration that compliance was accurately and objectively assessed
Participant training and monitoring	a. Description of training procedures b. Use of run-in or training periods c. Procedures to enhance compliance
Data management procedures	a. Discuss data management decisions that affect data analyses b. Define missing data criteria and actions
Data analysis	a. Rationale for aggregated or disaggregated approach b. Clearly specified model used in analyses c. Details of procedures, such as autocorrelation approach, random effect levels

The best way to obtain a representative sampling is to ask participants to record their pain at random moments throughout the day. A stratified random sampling plan, a variant of pure random sampling of times and one of the more common EMA schemes, can work nicely to accomplish this aim. Briefly, a stratified sampling protocol randomly samples moments within blocks of times (e.g., 2-hr blocks) running throughout the day (13). A preprogrammed device such as a palmtop computer can provide an auditory signal that alerts the participant to make a recording. This scheme ensures that the entire waking day is covered and that the moments sampled within blocks are random (although some small restrictions could be added to enhance the plan). Again, the rationale underlying these decisions should be clearly presented in the write-up.

There is an important implication of saying that the *entire* day should be sampled to adequately address this question: Random sampling should begin from the point an individual awakens and should end when he or she goes to bed. This design feature was very difficult to implement with the early devices for signaling diary entries, such as pagers and preprogrammed wristwatches, because there was no practical way to make the signals conform to participants' schedules. Even individualizing prompt schedules to participants' typical daily schedules is not adequate, because participants' schedules are prone to natural variation. These variations are not random in nature (e.g., an individual is stressed and awakes earlier than usual or a weekend party lasts into the early morning) and omitting them threatens the validity of the resulting summary measure. Study reports should state what degree of time coverage was afforded by the sampling scheme, the rationale for that coverage, and the implications of any limitations on coverage.

The decision about how many samples or recordings per day are needed should also be guided by the nature of the phenomenon to be recorded. Referring to our example, knowledge about the behavior of pain intensity over the course of the day—for instance, whether the pain is constant or episodic, the duration of an episode, and so forth—will inform the sampling density. Brief and infrequent episodes indicate that a high density of sampling is probably required, whereas longer episodes can be adequately assessed with a low density of sampling. It should also be noted that *very* brief episodes might never be captured by immediate momentary sampling, even when fairly dense sampling frequencies are employed. A variation of momentary reports that allows some retrospection might be necessary to capture such brief experiences.

Similarly, a daily sampling scheme also should be influenced by theoretical considerations. This is especially true when an investigator is interested in associations between two or more *within-subject* factors. In this case, a theoretical position about temporal relationships among the variables is critical for specifying the sampling scheme that provides an adequate test of the hypothesized associations. In our example, knowing the time lag from medication administration to pain assessment may be paramount for interpreting the pain data because pain relief is thought to decrease with increasing time from medication. In some instances, theory may justify a sampling scheme that oversamples certain times, based either on time of day or time since some event (e.g., 14). In any case, investigators should carefully consider how experience is sampled, taking into account the principles of representative sampling, their conception of the underlying phenomena, and the hypotheses being evaluated.

Having considered issues related to sampling, we next consider one of the most commonly used approaches to diary data: daily end-of-day (EOD) diaries. This approach has participants complete a diary at the end of the day, usually about the experiences and feelings of the entire day (e.g., assessing the day's mood, pain, or stressful events). Many researchers (including us) have used EOD study designs and have described the attractive

features of EOD sampling designs (15). EOD studies generally do not impose excessive burden on participants and are thus very useful for studying people over long periods of time (e.g., many weeks or several months). However, it is notable that EOD approaches do not adopt a sampling approach. The data are collected at end of day as a matter of convenience and rely on recall rather than momentary report. This is problematic when the constructs to be measured are prone to recall bias even within a day. For example, many EOD designs ask participants to rate their mood for the entire day. This may be a problem because we know that mood fluctuates considerably over the course of a day. The EOD mood summaries, which rely on recall and summary processes, may be subject to recency, peak, and summary biases.

However, if a researcher is interested in salient and relatively discrete occurrences, such as a simple count of migraine headaches or significant arguments with spouses, then EOD recall methods may adequately capture the desired information. Investigators considering EOD methods also should consider that the end of the day is not a representative timepoint for many subjective states (e.g., participants are likely to be tired, at home, not engaged in work). This adds to the potential for recall bias in EOD reports. In summary, careful thought must always go into the decision about the degree of retrospection acceptable considering whatever is known about the construct in question and about the rationale for selecting the timing of recording. The rationale for the decision should be reported in the introduction to the article.

Careful thought is needed to decide the number of days to be sampled and the arrangement of those days over time (e.g., consecutive days or separated by some interval over a longer period). One consideration in our example is the degree of stability of pain over days (16). Generally, there is no simple answer to the question, "How many samples per day should be taken and for how many days?" Increasing the number of samples each day clearly provides increased temporal resolution and statistical reliability, but it also increases the burden to participants. However, studying people over longer intervals (more days) may increase generalizability in terms of days and allow one to observe a greater range of response. Statistical power also is affected by the number of samples per day and the number of days in the protocol. However, how power is affected depends on the nature of the hypothesis being testing (e.g., between persons versus within persons) and the sources of variability in the outcome variables. Interested readers should refer to the work of Raudenbush (17) on this topic. In summary, considerations of sampling design do not lend themselves to formulaic answers but require careful weighing of theoretical and statistical, as well as practical, considerations.

Specify the Details of Sampling Procedures

Once the type of sampling scheme has been determined, a number of decisions must be made concerning the implementation of the scheme. Although this may at first glance appear to be a minor detail deserving only passing attention, many of the decisions have important implications for evaluating the validity of the assessment methodology.

Participant discretion in responding. A research report should describe any provision that allows participants to select when they record their data. When participants can select the timing of recording, bias may inadvertently be introduced (e.g., participants may remember to make reports of pain when they are feeling especially sharp pain or may find it more convenient to make reports when they are free of pain). If momentary assessments are to be made at particular times of day, the report should provide data about participants' reported or documented compliance with the prescribed times.

Missing and delayed data. For many reasons, participants will, on occasion, not respond to the signal to make a momentary recording. The signaling device (pager, watch, palmtop) may fail, or participants may not perceive the signal because the signaling device has been left at some distance from them or they may be in an environment where the signal is drowned out by other noises (e.g., a sporting event). Not much can be done in this case, and resulting data should be recorded as missing. A description of how signaling was implemented may be essential to understanding any biases that may have been introduced by the particulars of the signaling method.

In circumstances when the participant is unable to respond immediately to the prompt, the investigator has alternatives regarding the protocol. One alternative is to instruct the participant to simply ignore the prompt and resume data entry at the next signal. A second option is to allow the participant to complete the assessment some time later, but this option raises a question about what the target of the reporting should be: a recollection of the experience at the time of the missed signal or the experience at the (delayed) moment of recording? If the former, then we must ask ourselves how much elapsed time from the original signal is acceptable to consider the recollection valid? For example, some researchers are sanguine about recollections made within 5 to 10 min of the prompt (generally, we are) but would be uncomfortable with delays of 30 min or more, fearing systematic bias akin to those associated with longer recall periods. These decisions need to be made thoughtfully and should be clearly presented in research reports. It is desirable to present an analysis contrasting immediate and delayed responses to assess potential bias.

Programs for computerized EMA have been developed that allow participants to delay making an entry if they are unable to respond when signaled.[1] Some programs allow repeated delays that can result in an entry being made many minutes after the original signal. Although these entries may be viewed as valid by the investigators, it is incumbent upon investigators to inform the reader regarding definitions of an acceptable response. It is significant that early signaling methods (pagers and wristwatches) did not allow for this flexibility and did not provide information about when a response was actually recorded, a point we return to later.

[1]Some of these programs also allow the participant to specify a reason for the delay in response, which can be useful for interpreting the data.

Suspension of prompting. Another feature of sophisticated computer EMA solutions is the option to suspend signaling for a period of time when auditory signals would be embarrassing or inconvenient. The duration of this period may range from a few minutes to several hours, although participants also may have the capability to discontinue the suspension if their circumstances change. Applied judiciously, such provisions may help participants adapt a demanding protocol to their daily lives. If overused, such provisions can undermine a well-conceived momentary sampling plan, and the frequency and duration of such suspensions of sampling should be reported so that readers can make their own appraisal of the data.

Describe the Data Acquisition Interface

Successful EMA applications are based on many factors, including user-friendliness, simplicity of design, and thoughtful uses of response interfaces. By the latter, we mean the various ways that participants can indicate their responses on either paper diaries or the screens of palmtop devices. Of course, the concerns we discuss apply to the design of any questionnaire, but EMA applications present two additional issues. First, participants will be assessed frequently, usually many times a day. This means that each sampling occasion must be relatively brief (e.g., 1 to 3 min). Otherwise, participants may quickly become annoyed with the task, which results in poor compliance (assessments either not completed or completed sloppily) and dissatisfaction with the study. Care in the design of questions and response formats can greatly reduce the burden by enabling minimal response time and by making the task more pleasant. To date, this important practical issue has not received much attention in the momentary assessment literature. Research reports should communicate how questions were presented and how responses were gathered.

The second set of concerns is based on the unique formatting demands that EMA diaries have. So that they may be easily carried throughout the day, EMA diaries use very small formats, ranging from pocket-sized booklets (for written diaries) to small palmtop computer screens (such as personal digital assistants). This small format means that compared with typical questionnaires, relatively little real estate is available for questions and response keys. Because questions are often presented in a telegraphic format to conserve space, care must be taken to ensure that the original meaning of the full question (and responses) are preserved in the momentary application. Additionally, because EMA items are presented multiple times, brevity and clarity are important virtues: Participants simply stop "seeing" items that are long or verbose. We recommend that investigators consider the optimal display of items for an EMA application and routinely report the text of their primary EMA questions and any procedures undertaken to "translate" items from their original versions and to validate the translation.

A corollary of this issue arises from the flexibility that is possible with palmtop computers. Many palmtop applications run real-time algorithms that alter their behavior according to predefined parameters set by the programmer. For example, in one of our own studies, the sampling rate increased when participants indicated that they were experiencing moderate or high levels of psychological stress (18). Other investigators have used similar branching schemes to change the question content presented to participants based on responses to earlier questions (e.g., skip-out routines). Although this plasticity is welcome and empowering to creative research protocols, it creates a level of complexity previously unknown to the field.[2] For this reason, we believe it is necessary for investigators to report on the algorithm employed for computer-based EMA applications in the Methods section of the resulting paper. Only the essential principles need be reported to allow other investigators to replicate the study.

Report Compliance with the Sampling Plan

Success of an EMA study depends on a high degree of participant compliance with the sampling scheme protocol; the validity of the assessment scheme is threatened by noncompliance. Using the pain example presented earlier, if participants did not respond to a high proportion of prompts, say 20% or more, the investigator might question the representativeness of the sampling[3] and, thus, the validity and generalizability of the findings. Random noncompliance is certainly more acceptable than systematic noncompliance, but we believe that most noncompliance is systematic rather than random. Using pain as an example, it is easy to imagine ways that pain intensity could systematically influence responding. When in considerable pain, participants might choose to ignore a prompt because it is too overwhelming given their current condition; in doing so, the resulting average level of pain for the period studied would be falsely underestimated (the variability of momentary pain scores would be reduced as well). Alternatively, participants may not complete a signaled assessment when free of pain, mistakenly believing that the researchers are only interested in reports when the patient is in pain. The effect of this response behavior would be to increase overall pain summary scores because pain-free intervals would be systematically underrepresented. Effective training in the momentary protocol, discussed later, may largely eliminate this type of response bias. Our point, though, is that missing data are likely to be systematic and may be a threat to a study's validity.

A related issue involves the possibility that participants are not completing momentary reports according to the protocol but are completing reports at a later time. An extreme example of this problem occurs when an entire day's reports are hoarded and completed at the end of the day or, even worse, days later. Furthermore, participants may indicate that they have completed the reports according to the protocol schedule when in truth they completed the reports at another time. This deception may at first glance appear unlikely, but recent studies of patient compliance with medication regimens suggest this occurs fre-

[2]This complexity also increases the probability of errors in programming that may be especially hard to detect in field applications. Extensive pretesting and validation of EMA computer programs is essential.

[3]This criterion level of response is arbitrary, and an argument could be made that it should be considerably more stringent depending on the reasons for the missing data.

quently (19). For instance, studies using aerosol inhalers outfitted with electronics to monitor when a patient used the device have shown quite dramatic instances of hoarding wherein many doses of medication were logged by patients immediately before a scheduled appointment with their physician (20).

Measurement of compliance is difficult when responses are recorded on paper. The investigator knows when the recordings were *to be made* and what date and time *were entered* on the record form, but there is no way of knowing when responses were actually recorded. Some investigators have developed procedures wherein diaries are returned to the investigator on the day following completion, allowing the investigator to check postmarks. At best, this procedure only ensures that responses have not been delayed longer than 24 hr. However, participants could still complete diaries the day after they should have been completed and still meet the study postmark criteria. Other investigators have developed ingenious schemes wherein a programmed wristwatch provides a unique numerical code at the time of the signal and the code is to be recorded on the diary. This procedure probably improves compliance given the message it explicitly imparts to participants about compliance, but participants may defeat the plan by jotting down the codes and entering them later when the entire diary is completed. Of concern is that there have been some informal reports of this happening based on participant debriefing at the close of studies.

We recently completed a study that confirmed our concerns about compliance with paper-and-pencil diaries (21). Using a specially instrumented paper diary binder that enabled us to covertly monitor the openings and closings of the diary (with a computer embedded in the diary), we were able to document actual timely completion of diaries and compare it with the time and date participants entered on the written diary cards. Across a 3-week reporting period, participants submitted cards corresponding to 90% of the assessment occasions. However, only 11% of entries were actually compliant (based on the diary binder being open and having a completed diary card for the corresponding time). Moreover, there were indications of frequent hoarding of diaries across days. This result suggests that poor compliance is a major issue in research using paper diaries.

A solution to accurate recording of compliance is to have a mechanism wherein each diary completion is time and date stamped in real time in a manner that is not accessible to participants. This has been implemented in several palmtop EMA solutions, and we strongly recommend this approach. Without such procedures, the investigator simply will not know the true degree of compliance. However, if faced with the absolute necessity of having to run a protocol without objective time and date stamping, we recommend that a thorough debriefing of participants be included in the protocol because some studies have shown that a portion of noncompliant individuals readily admit to their noncompliance and, somewhat surprisingly, to their faking of timely reports (19). It is essential that methods for ascertaining compliance be included in scientific reports along with the resulting data.

Beyond assessing and recording compliance, investigators should do what they can to achieve high compliance. This often requires a combination of participant training, monitoring and feedback, incentives, and other procedures (22). For example, the study reported with paper diaries also examined compliance rates with a similar protocol using an electronic diary and compliance-enhancing procedures. The electronic diary system achieved 94% compliance.

We have four recommendations concerning reporting of compliance. First, investigators should always report the missing data rates in detail. We have seen momentary studies in which no mention is made of missing data, implying the improbable situation in which the completion rate was 100%. In addition to reporting the average number of missed assessments, the report also should characterize the range. Analyses should consider whether participants' missing data rates are associated with the time of day. Second, the "rules" an investigator uses for designating a diary or a participant as valid or invalid should be explicitly stated and justified. The number of participants excluded on these grounds should be reported, and attention should be paid to the potential for group differences and attendant bias. Again, this will provide consumers of the research with adequate information to replicate the data management scheme employed in a particular study and to assess the validity of the reported findings. Third, compliance should be assessed against the original participant protocol. Specifically, investigators should report the percentage of study assessments that were missed. The definition of compliance should correspond exactly to the original protocol specification rather than define some new (more liberal) standard for compliance (e.g., 23). Fourth, we believe it is now incumbent on researchers to demonstrate that their compliance rates are accurate, given the data we presented from the instrumented diary study.

Report the Procedures Used to Train and Monitor Participants

EMA studies usually place considerable demands on participants. They must be willing to complete repeated assessments in a timely way. In sampling protocols, they must be willing to be signaled throughout their waking days, must carry materials with them at all times, and must be willing to devote considerable time and effort to a project. Therefore, participants in momentary studies require thorough knowledge of what is expected of them and thorough training in the materials they will be using. Effective training improves compliance and increases the likelihood that procedures are followed correctly, especially when participants are viewed as partners in the research endeavor.

Some investigators have opted for extensive training on procedures, including training on the meaning of items and the available alternative responses. The return on investment in training is likely to be favorable because participants complete the assessment many times. Documentation also may be provided in the assessment itself, whether on paper or on computer. Such training should be documented in the research report.

To date, we have seen only sparse descriptions of training methods in the Procedure sections of papers employing momentary methods. Nevertheless, we know that the range of training experiences offered by EMA investigators is broad. Some investi-

gators employ rather causal, individualized training, whereas others employ systematic, comprehensive training involving demonstrations of momentary data collection and participant practice with feedback during the training. Some investigators extend training into the field by having participants use the diary for several days and then return for feedback about their data. In some cases, feedback continues periodically throughout the study. We strongly suspect that the thoroughness of the training and feedback given to participants affects compliance and the validity of the data. It is difficult to interpret the missing data rates (if reported) without knowledge of the training methods. Just as important, the field will not advance unless investigators are able to replicate each other's findings; this is impossible to do accurately without the provision of training details. Training may be especially important for protocols using palmtop computers because some participants may not be familiar with these devices. In any case, procedures for training and for providing technical support and feedback (if any) to participants should be reported.

Report on the Data Management Procedures

One of the surprising discoveries that confronts a first-time momentary investigator is the mountain of data these techniques generate. A relatively simple protocol with 50 participants reporting on their pain, affect, and environmental settings (e.g., activity, place) eight times a day for 1 week will generate more than 50,000 data points (assuming 20 data items per momentary assessment). If the protocol is implemented on a palmtop computer running a sophisticated EMA program, then the number of data points may increase dramatically because there are many additional records generated each day concerning other transactions and compliance-related data. More complex protocols generate much more data.

The investigator must design a plan for checking the data to ensure that they are valid and should describe data-handling methods well enough that other investigators know exactly what was done. Data management procedures are needed in all empirical investigations, but data from EMA studies require many decisions not faced by studies that yield less complex data sets. A couple of examples provide the reader with a sense of these decisions. In the pain study described earlier, let us imagine that a participant went to a party and did not go to sleep until 2 a.m. Should the pain reports that occurred after midnight be included as part of the 1st or 2nd day? This decision may have important implications for day-of-the-week analyses, for instance. Another example concerns how a partially completed diary should be handled. What proportion of data is required to count a set of entries as missing? There are a host of decisions along these lines that experienced EMA investigators are all too familiar with. In the spirit of full disclosure, we suggest that investigators carefully consider which of their decisions have implications for the analyses and report those in their papers.

Describe the Data Analysis Approach

The volume of data referred to earlier must not only be managed but also be analyzed. As with all large, complex data sets, an a priori plan of attack that sets the course for analyses is incredibly important, lest the researcher become caught up in pursuing incidental analyses. Because analysis of EMA data is not a settled area, descriptions of the precise methods and statistical procedures that were used for analyzing the data and testing hypotheses should be reported. Many standard statistical techniques (e.g., repeated measures analysis of variance, ordinary least squares regression) are usually appropriate methods of analysis if the momentary data have been aggregated to the subject level. Even so, aggregate momentary data present unusual, though manageable, issues that must be addressed in the report (e.g., whether the number of observations in the aggregate varies across participants and leads to differential reliability). As we and others have previously discussed (24–27), different methods are needed when the momentary data are not aggregated. Investigators should discuss the rationale for selection of aggregate or disaggregate analysis and report them in enough detail so that other investigators are able to reproduce their analytic methods. Because different modes of analysis are available, care must be taken that the analytic approach matches the hypothesis being pursued. Again, experienced EMA investigators are familiar with the vagaries of these complex analytic procedures, which demand not only technical expertise but also experience with complex statistics and unusual data structures (e.g., such as when a maximum likelihood procedure converges to a local minimum or choosing the particular autocorrelation structure that best fits the observations).

CONCLUSION

We recognize there are no hard and fast rules about conducting studies and publishing their results in any area of research. Nevertheless, given the burgeoning interest in EMA approaches to self-reported phenomena and our own experience with these methods, as well as the fact that EMA studies require particularly careful thought in conception, design, implementation, analysis, and reporting, we think it reasonable to propose these general guidelines. We also wish to acknowledge there are issues we have not discussed that are nonetheless important for this research area. As a starting point, however, we believe it is essential for EMA studies to adopt and convey a systematic approach to sampling and assessment and for papers reporting EMA studies to contain standard and detailed accounts of the methods and methodological results. This approach will allow readers to better interpret EMA findings and, over time, will allow the field to assess the suitability of diverse methodological approaches. It is our hope that consideration of the factors discussed here will better enable researchers to make informed decisions about their momentary study designs and will result in more complete and useful published reports about their efforts.

REFERENCES

(1) Stone A, Neale J, Cox D, et al.: Daily events are associated with a secretory response to an oral antigen in humans. *Health Psychology.* 1994, *13:*440–446.

(2) DeVries M: *The Experience of Psychopathology.* Cambridge, England: Cambridge University Press, 1992.

(3) Stone A, Turkkan J, Jobe J, et al.: *The Science of Self Report.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2000.
(4) Bradburn N, Rips L, Shevell S: Answering autobiographical questions: The impact of memory and inference on surveys. *Science.* 1987, *236:*151–167.
(5) Gorin AA, Stone AA: Recall biases and cognitive errors in retrospective self-reports: A call for momentary assessments. In Baum A, Revenson T, Singer J (eds), *Handbook of Health Psychology.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc. (in press), 405–413.
(6) Erskine A, Morley S, Pearce S: Memory for pain: A review. *Pain.* 1990, *41:*255–265.
(7) Salovey P, Sieber W, Jobe J, Willis G: The recall of physical pain. In Schwarz N, Sudman S (eds), *Autobiographical Memory and the Validity of Retrospective Reports.* New York: Springer-Verlag, 1993, 89–106.
(8) Schwarz N: Self-reports. How the questions shape the answers. *American Psychologist.* 1999, *54:*93–105.
(9) Schwarz N, Sudman S: *Autobiographical Memory and the Validity of Retrospective Reports.* New York: Springer-Verlag, 1994.
(10) Salovey P, Smith A, Turk D, Jobe J, Willis G: The accuracy of memory for pain: Not so bad most of the time. *American Pain Society Journal.* 1993, *2:*181–191.
(11) Stone A, Broderick J, Porter L, Kaell A: The experience of rheumatoid arthritis pain and fatigue: Examining momentary reports and correlates over one week. *Arthritis Care and Research.* 1997, *10:*185–193.
(12) Reis H, Gable S: Event-sampling and other methods for studying everyday experience. In Reis H, Judd C (eds), *Handbook of Research Methods in Social and Personality Psychology.* Cambridge, England: Cambridge University Press, 2000, 190–223.
(13) Delespaul P: *Assessing Schizophrenia in Daily Life—The Experience Sampling Method.* Maastricht, The Netherlands: Maastricht University Press, 1995.
(14) Shiffman S: Real-time self-report of momentary states in the natural environment: Computerized Ecological Momentary Assessment. In Stone A, Turkkan J, Jobe J, et al. (eds), *The Science of Self-Report: Implications for Research and Practice.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2000, 277–296.
(15) Stone A, Valdimarsdottir H, Katkin E, et al.: Stressed-induced autonomic reactions are associated with altered mitogen responses. *Psychosomatic Medicine.* 1991, *2:*211–212.
(16) Stone A, Neale J: Development of a methodology for assessing daily experiences. In *Advances in Environmental Psychology: Environment and Health* (Vol. 4). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., 1982, 49–83.
(17) Raudenbush S: Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods.* 1997, *2:*173–185.
(18) Marco C, Schwartz J, Neale J, Shiffman S, Stone A: Do appraisals of daily problems and how they are coped with moderate mood in everyday life? *Journal of Consulting and Clinical Psychology.* 1999, *67:*755–764.
(19) Litt M, Cooney N, Morse P: Ecological Momentary Assessment (EMA) with treated alcoholics: Methodological problems and potential solutions. *Health Psychology.* 1998, *17:*48–52.
(20) Rand C, Nides M, Cowles M, Wise R, Connett J: Long-term metered-dose inhaler adherence in a clinical trial. *American Journal of Respiratory & Critical Care Medicine.* 1995, *152:*580–588.
(21) Stone AA, Shiffman SS, Schwartz JE, Hufford M, Broderick JB: Patient non-compliance with paper diaries. *British Medical Journal.* 2002, *324:*1193–1194.
(22) Shiffman S, Hufford M, Paty J: Subject Experience Diaries in Clinical Research. Part 1: The patient experience movement. *Applied Clinical Trials.* 2001, *10:*46–56.
(23) Raymond S, Ross R: Electronic subject diaries in clinical trials. *Applied Clinical Trials.* 2000, *9:*48–58.
(24) Affleck G, Tennen H, Keefe F, et al.: Everyday life with osteoarthritis or rheumatoid arthritis: Independent effects of disease and gender on daily pain, mood, and coping. *Pain.* 1999, *83:*601–609.
(25) Michela J: Within-person correlational design. In Hendrick C, Clark M (eds), *Review of Personality and Social Psychology Series.*1990, 279–311.
(26) Schwartz J, Stone A: Data analysis for EMA studies. *Health Psychology.* 1998, *17:*6–16.
(27) West S, Hepworth J: Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality.* 1991, *59:*609–662.