Mobile Assessment Guide for Research in Schizophrenia and Severe Mental Disorders

David Kimhy¹, Inez Myin-Germeys², Jasper Palmier-Claus³, and Joel Swendsen^{4,*}

¹Division of Cognitive Neuroscience, Department of Psychiatry, Columbia University, New York, NY; ²Division of Mental Health, School of Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands; ³School of Community Based Medicine, the University of Manchester, Manchester, UK; ⁴University of Bordeaux, National Center For Scientific Research (CNRS) UMR 5287, 146 rue Léo Saignat, 33076 Bordeaux, France

*To whom correspondence should be addressed; tel: (33)05-57-57-10-10, fax: (33)05-56-90-14-21, e-mail: Joel.Swendsen@u-bordeaux2.fr

Mobile assessment techniques have been used for nearly 3 decades in mental health research, including in investigations of individuals with schizophrenia and other severe disorders. This article reviews the benefits of these data collection strategies relative to traditional self-report or clinician-administered measures administered in hospital or laboratory settings. A detailed discussion of the technical decisions facing researchers in the field is then presented, covering study design issues, questionnaire content development, and choices in hardware and software selection. Following these points, sample recruitment and retention strategies are discussed, as well as the main statistical issues that are necessary to consider in the exploitation of repeated measures data generated by this methodology.

Key words: methodology/ESM/EMA

Introduction

The capacity of mobile assessment techniques to improve both research and clinical interventions is increasingly recognized by major health foundations around the world, including the National Institutes of Health.¹ Of all health care domains, clinical psychology and psychiatry have been the most active in the use of these methods in order to overcome 2 longstanding barriers in understanding the expression of mental disorders. The first major barrier concerns the important differences between the natural phenomena under study and the methods used in their investigation. This discrepancy is most visible for the assessment of temporal relationships among variables. That is, the expression of many forms of psychopathology are characterized by a relatively short “life cycle” concerning the period of time in which a specific vulnerability or risk factor may influence the severity of

symptoms or the onset of abnormal behavior. Such phenomena are observable over periods that are typically limited to a matter of minutes or hours, while most standard methodologies apply assessments spanning weeks, months, or years. For this reason, traditional self-report and clinician-administered measures that assess global experience or syndrome status rely heavily on retrospective recall. They are therefore unable to examine important but rapidly changing phenomena such as stress reactivity, momentary cognitions or changes in affect, negative and positive reinforcement of abnormal behavior, and many other processes that are temporally proximal to the expression of symptoms.

An additional major impediment to understanding mental disorder etiology or symptom expression concerns the ecological validity of the existing literature. For example, laboratory protocols may confirm that alcohol or benzodiazepines block panic attacks in individuals with panic disorder, but this does not demonstrate that the individual would choose to use these substances as a means of self-medication nor that this motivation to use substances would necessarily block attacks in situations that cannot be reproduced in the laboratory. In the same way, the cognitive biases frequently associated with major depression express themselves by influencing the manner in which the individual interprets the numerous events and experiences that occur throughout the day. A clearer description of these momentary phenomena in vivo would permit a better understanding of how diverse vulnerabilities for complex mental disorders influence the emergence or exacerbation of psychopathology. Drawing from the experience of researchers in the field, the objective of this article is to present an expert review of the conceptual and methodological issues to consider in using mobile technologies to investigate schizophrenia and other severe mental disorders.

Mobile Assessment in Mental Health Research

Mobile data collection strategies have been used for nearly 3 decades in mental health research in the goal of overcoming these traditional barriers of time and context. Pioneering work by researchers in the 1980s defined ambulatory methods such as the “Experience Sampling Method” or “ESM”² (also referred to as “Ecological Momentary Assessment” or “EMA”³) as well as provided their first applications to clinical psychiatry.⁴ Much of this early work was exploratory or descriptive in nature, such as time-budget studies of the frequency of patient behaviors and activities. The application of mobile data collection to testing theories of etiology rapidly followed over the next decade, and these methods are now increasingly explored as a means of delivering interventions.

Concerning technical characteristics, almost all studies conducted in the 1980s and 1990s used paper-based methods where a preprogrammed wristwatch or beeper would signal the patient to complete an assessment form describing their experiences or behavior. These techniques have been applied with success in the study of a wide variety of psychiatric problems including personality disorders, mood disorders, substance abuse, and psychosis.^{5–12} However, paper-based methods have been progressively replaced by computerized assessments that use newer electronic technologies (eg, personal digital assistant [PDA] micro-computers or smartphones). While both approaches use mobile devices to indicate to subjects the moments throughout the day to provide specific information relative to their behaviors and other data, it is nonetheless important to note that electronic methods have several advantages over paper-based protocols. A principal benefit concerns the capacity of electronic methods to furnish accurate information on the timing of assessments rather than relying on patient estimates of the moment that data were collected. Such information is at times crucial for confirming the directionality of correlated variables. Electronic methods also allow for the branching of questions administered to patients, whereby each response may trigger subsequent questions as appropriate, and they provide considerable ease in data management. Additional benefits include confidentiality of responses that cannot be viewed by unauthorized individuals, and the facility at which ESM/EMA data may be coupled with other physiological measures such as heart rate, cortisol levels, or physical activity. The use of these “second generation” mobile has recently been validated in patients with schizophrenia, including hospitalized¹³ and outpatient¹⁴ samples, demonstrating good compliance rates and high concurrent validity with traditional clinic-based measures. This latter issue is particularly complex, however, as it necessarily raises the question of which measures should be used as the “gold standard” by which other instruments are compared and evaluated.

Mobile Assessment Study Design

Sampling Strategies

This section reviews the main sampling strategies and issues to consider in the construction of electronic questionnaires, as well as the advantages and disadvantages of the different options available to investigators.

Event vs Time Sampling. An important initial decision for the mobile assessment researcher is the manner in which information will be collected concerning daily life experiences. Event-based sampling constitutes one option whereby the occurrence of a specific event (such as having a social interaction, smoking a cigarette, or having an argument with a spouse) indicates that a report should be completed. Event-based sampling seems intuitively useful because the phenomena of interest are definitely covered by the study, and it is well adapted to assessing relatively rare or salient events that may not occur everyday. However, this option is also a closed observational system, including information regarding only that particular moment of assessment. In addition, concerns have been raised that event sampling may potentially induce changes in behavior, such as that people may be discouraged from performing certain behaviors targeted by the study just to avoid writing another report.¹⁵ Importantly, event sampling also imposes selection decisions upon the participant such as deciding if just saying “hello” constitutes a real social interaction, if one puff of a cigarette constitutes “smoking a cigarette,” etc, and therefore induces doubt if reporting certain events is necessary. Relying on participants to make decisions of this type may hamper the compliance and reliability of the method. By contrast, a time-sampling approach, with data collection being dependent on a time schedule rather than on events, is an open observational system covering different situations and contexts (including target and nontarget moments). In addition, no selection decisions are required by the participant, which should improve reliability and provide more accurate estimates of event frequency.

Random vs Fixed Time Sampling. For investigations using time sampling, an important supplemental decision concerns whether assessments should be administered at random or fixed intervals of time. Although this issue has been the subject of extensive debate among ESM/EMA researchers, there is very little empirical data to date demonstrating the superiority of either strategy. Investigators must therefore decide on the appropriateness of a given approach based on the scientific objectives of a particular study. Sampling at fixed time points with regular periodicity has clear advantages in terms of statistical modeling, such as time series analyses that have been developed initially to deal with data collected at stable time intervals. Fixed time points may also be more appropriate for

examining certain research questions, such as the study of routines (assessed at the same moment across days) or events that are linked to a particular time of day (eg, end of work). However, there are also several disadvantages to fixed schedules. First, data in a fixed sampling design do not represent the full daily behavior patterns of participants, in particular if the report examines behavior only at the moment of the assessment. Using this approach, a specific number of moments have a 100% probability of being selected, whereas the rest of the moments over the day are never selected. In choosing random sampling, by contrast, each moment has an equal probability of being selected. It is unpredictable, and therefore provides a better global estimate of how people spend their days (in particular if assessments target behaviors or experiences at the moment of the report). In practice, most studies using random assessments divide the day into blocks of approximately 90 minutes, and a random moment is selected within each block in order to provide a stable pattern over the day.¹⁶ Random reports are therefore preferable for examining scientific questions such as time budgeting of momentary behaviors and experiences. However, just as some investigations may include a mix of event- and time-based assessments, many research protocols represent a compromise between random and fixed assessments by asking participants to report behaviors or experiences spanning the minutes or hours since the last assessment. Protocols of this type theoretically reduce differences between these 2 assessment choices, but it remains to be investigated the degree to which retrospective memory biases may reduce the validity of such data.

A frequently raised concern is that the structure of fixed time protocols may be easily recognized by participants and therefore might influence subject behavior in anticipation of the assessment (eg, a participant may stay home in order to complete the assessment in a quiet environment). Researchers may again debate the magnitude of this effect, with some arguing that investigations of psychological theories have shown essentially identical results whether fixed or random assessments are used,^{10,11} while others have argued that knowledge of assessment timing may, at least theoretically, increase the reactivity to the method.¹⁵ Only a direct comparison between these 2 approaches would provide the answer to this enduring debate, as well as concerning other details of study designs that constitute a source of disagreement among daily life researchers.

Number of Days and Number of Assessments. The selection of number of days and number of assessments is dependent both on the expected periodicity and time course of the processes under investigation, as well as on feasibility. Overall, it is important to have a balance between number of sampling days and number of reports each day. While a wide range of time points has been used, ranging from

1 day to several weeks,⁸ as a rule of thumb it may be wise to include at least 6 days because this inevitably includes week and weekend assessments (and therefore provides a more global characterization of daily life experiences).

In considering the number of assessments per day, missing data is expected in investigations using mobile technologies and should be anticipated in choosing the assessment frequency. For some research questions, it may therefore be a good strategy to increase the sampling load in order to assure enough data for statistical analyses (especially if time-lagged strategies are used in which data from previous moments predict current state). However, a greater frequency of assessments per day may not necessarily be experienced as more intrusive or burdensome by the participants. For example, it may be argued that when a participant misses 1 or 2 assessments out of a total of 4 that were scheduled, they may become concerned about their compliance, question if the device still working or wonder if they should contact the researcher. With a higher frequency of assessments, it may more easily become part of their daily routine and missed assessments will be more easily normalized.

Questionnaire Development

Constructing Questionnaires. One of the most crucial steps of mobile assessment research is the development of the questionnaire,¹⁷ as momentary data collection is very different in nature from standard cross-sectional questionnaires. Whereas cross-sectional questionnaires are based on global retrospective recollections, often cued by an extensive narrative or metaphor, momentary questionnaires should consist of self-exploratory questions with short cues that reflect upon commonplace situations. Items from cross-sectional questionnaires may be a good starting point, but one has to be careful that items really reflect momentary states rather than traits (eg, putting “right now” in front of a trait-like item does not necessarily make it a momentary item).

When constructing a questionnaire, it is also important to use language that reflects how people describe their own behavior and experiences. Psychological vocabulary that lay people are not familiar with (such as “attributions,” “coping,” or “dissociation”) should be avoided.¹⁷ In addition, the frequency and specificity of items need to be kept in mind. Extreme items may pertain to fewer situations and show less variability compared with more mildly formulated items. Negatively formulated items are less frequently endorsed, leading to skewed distributions, whereas positively formulated items usually have a normal distribution. Finally, it is important to avoid reflective questions that link 2 distinct constructs, such as “in this social context, I feel down.” Much more information can be gained when mood and context are asked independently and when the association is made

statistically by the researcher after data collection. The independence of questions would also shed more light on patterns that participants are not consciously aware of, while also limiting the risk of socially desirable responses.¹⁸

As previously noted, questionnaires often consist of a mix of momentary items and questions that span the period since the last assessment. It is possible that current mood states may potentially influence recall or interpretation of previous experiences. When creating a questionnaire, it is therefore suggested to order it so that the most transient experiences such as thoughts, mood, or symptoms are at the beginning of the questionnaire, followed by more stable items such as context, and to present the retrospective items in last position.¹⁷ Different item formats have been used, including bipolar scales (eg, -3 to 3), visual analog scales and Likert scales (eg, 1-7). While there is no evidence for absolute benefits of a particular item format, it is important to remain consistent throughout the questionnaire. Open-ended questions provide an excellent opportunity to focus participants in the moment (eg, "What are you doing at the moment?"). However, they are time consuming and often difficult to implement in electronic assessments. In the latter case, category boxes are often used to collect contextual data. Participants should be focused in the moment, then inquiries about the context can be made (eg, "I like this activity") before assigning it to a category in order to avoid global or vague answers.

Because participants have to answer questions on a frequent basis in their normal daily lives, the duration of the assessment should generally be limited to between 1 and 3 minutes. This limit would nonetheless permit a large number of open- and closed-ended questions. However, in designing the questionnaire, a frequent concern is that the repeated assessment of mood or symptoms may influence the person's experience (and potentially exacerbate symptom intensity). One useful strategy to avoid this influence is to intersperse-specific items with more general items in order for individuals not to be solely provided with emotionally salient information. For example, assessing delusions while also including questions on the context of assessment may lessen the otherwise heavy focus on symptoms. In the same way, assessment of negatively valenced states (anxiety and sadness) may be accompanied by questions concerning positive or neutral states (happy mood and energy level). In this case, a larger number of items is preferable. Alternatively, if the goal was to change behavioral patterns, for example in a therapeutic intervention, a minimal number of items would be preferable in order to improve the learning curve and keep the patient focused on the therapeutic objective.^{19,20}

A frequent question when using repeated daily life assessments concerns the possibility that participants may lose motivation and that missing data would therefore increase over time (fatigue effects) or that repeatedly asking an individual how they think, feel, or behave may

change the intensity or frequency of those variables (reactive effects). Investigations that have directly examined these potential biases have found no indication of fatigue or significant reactivity to mobile assessments in patients with schizophrenia.^{14,21} However, despite the brief duration of each electronic interview, it is possible that participants may still become irritated or tired, and compliance with the methodology may become very inconvenient at certain moments. It may therefore be desirable to include an item inquiring about these undesirable effects so that it can be used as a covariate in the analyses.

Devices and Software

Perhaps, the most basic decision in mobile assessment research is choosing whether to use a paper-and-pencil strategy or computerized devices. Both have strengths and weaknesses that should be considered relative to the needs of a given study. Paper-and-pencil approaches are inexpensive, easily available, and user-friendly for the great majority of participants. As previously noted, however, a major disadvantage of this approach is that the researcher cannot be sure about exact time point of data entry, possibly hiding significant compliance problems. Although past research has shown that compliance rates are comparable for both strategies,²² data entry is time-consuming and branching or linking new questions to specific responses is not possible. Computerized mobile assessment overcomes these constraints and its use is rapidly growing in investigations of schizophrenia as well as in psychiatry in general.^{13,14,21} Advantages include easy data input, exact information on response times, possibility for branching, and increased speed of assessment completion. The respective disadvantages of this second generation of ESM/EMA methods include technical problems (eg, battery problems, broken screens, software issues), "user friendliness" (especially for computer-unfamiliar participants), and the difficulty of including open-ended questions.

The assessment device should be small, reliable, cost-efficient, and flexible (working in all, or at least most, circumstances) in order to increase compliance with the method.¹⁵ The device should also preferentially not offer too many distractions (eg, games) in order to keep subjects focused on the study but also to decrease the risk that the device is stolen or that the battery dies quicker. As for hardware options, paper-and-pencil studies have often used preprogrammable watches (eg, Timex Ironman watch). These watches are inexpensive and easy to program, with a long battery life. For computerized approaches, most studies have used PDAs and smartphones. Because these devices have become more affordable over the years, computerized mobile assessment has become a valuable option. However, device selection should be informed by length of planned assessment (number of days) and by mobile device battery life.

For protocols of longer duration, researchers may need to provide participants with battery charging equipment. In addition, for studies conducted on inpatient units, researchers may want to consider selecting mobile devices without video or audio recording capabilities to protect of the confidentiality of other patients on the unit or to lock out these functions.

Several software packages, such as the Purdue Momentary Assessment Tool,²³ MyExperience,²⁴ and ClinTouch²⁵ (as well as several home-made programs) have been used for programming experience sampling protocols on these devices. Recently, a new device, PsyMate, has been developed specifically for mobile assessment research²⁰ and that has the advantages of high user friendliness, easy programming, and potential for real-time interactive applications. Information technology (IT) support may still be necessary in the preparation of devices for the study, and some forms of data collection (such as text messaging) may not be feasible because one is not sure about the actual time of transmission. Many existing software programs provide the option of locking out the device from all functions other than those required for the study in order to conserve battery life or offer other options for participants such as delaying responding if the assessment occurs at an inconvenient time. However, these options remain the choice of the researcher and do not constitute strategies for which any consensus has been achieved.

Practical Guidelines for Sample Recruitment and Study Management

Sample Recruitment

Clinical Considerations. With the notable exception of individuals with severe disorganization preventing them from effectively completing assessments or those who are heavily sedated from medications, studies of schizophrenia spectrum disorders have covered a wide range of symptoms and clinical profiles. These studies have successfully included individuals with acute paranoia, active auditory and/or visual hallucinations, delusions of control, reference, and grandiosity, as well as concurrent alcohol or drug use.^{26–30} The sample sources have included both treatment-seeking^{13,29,31} and nonclinical settings^{10–12} and both paper-based^{26,28} as well as computerized assessments using mobile devices were successfully used.^{13,14,32} These latter reports also indicate that few participants discontinue their participation due to exacerbation of clinical symptoms. For these reasons, the recruitment of individuals spanning the full spectrum of psychotic disorder severity into mobile assessment studies does not present unique challenges beyond those typically present in more classic study designs of such populations. Furthermore, the expansion of applications for mobile assessment in these populations appears highly feasible for both research and treatment. For example, Kimhy *et al.*³² recently

completed a study of hospitalized patients with psychosis in which monitoring with mobile devices was integrated with concurrent measurement of ambulatory cardiac autonomic regulation. In other studies, mobile devices were successfully integrated into cognitive-behavioral therapy (CBT) interventions^{31,33} demonstrating the particular flexibility of this approach.

Considerations for Inpatient and Outpatient Samples. Despite the successful application of mobile assessment in schizophrenia research, alterations or adaptations to the sampling protocol or procedures are often needed in order to accommodate needs of a given sample. Concerning research with inpatients, an obvious first element for successful execution of the study concerns cooperation of the unit's clinical staff. Some researchers have found it helpful to conduct a brief presentation of the study to the unit's personnel describing its purpose and procedures, as well as troubleshooting strategies, prior to beginning the study. Depending on the assessment schedule, it may also be helpful to provide information to the evening/night staff concerning which patients are participating to minimize potential problems. Additionally, in scheduling mobile assessments, it is beneficial to consult with the unit's clinical staff regarding other appointments the patient may have scheduled outside the unit (eg, functional Magnetic Resonance Imaging [fMRI] assessment) that may potentially interfere with the research protocol, as well as regarding the patients' schedule of clinical activities on the unit (eg, group therapy). For example, Kimhy *et al.*^{13,32} successfully addressed this latter issue by instructing participants (in agreement with the unit's clinical staff) to excuse themselves from the room where the clinical activity is being held for a few minutes in order to complete mobile assessments without creating a disturbance and then to return to the activity a few minutes later. Similar arrangements were also successful for individual appointments with clinical staff in the same unit. In terms of outpatient or community-based samples, there are obviously fewer constraints for conducting the study around hospital schedules and appointments. However, outpatient samples require careful planning concerning return of the mobile device, which should ideally follow the end of the sampling protocol as closely as possible in order to avoid its loss or damage. As there are no clinical staff on hand to remind outpatients to recharge the device or to keep in with them at all times, careful training and supervision of such samples is necessary to decrease data loss or compliance issues.

Sample Characteristics and Study Compliance

Age and Gender Considerations. While younger participants may have better familiarity with PDAs or smartphones, age does not appear to play a significant role in participants' ability to successfully use such devices. For example, Granholm *et al.*¹⁴ completed a study of patients with schizophrenia living in the community in

which the average age of participants was 44.06 years and with a significant portion of older individuals ($SD = 10.46$). Similarly, while some gender differences in specific domains of functioning have been documented in mobile assessment studies, there are no reports of gender differences in compliance rates.

Language. Participants' demonstration of familiarity and ease with the vocabulary used in which the questionnaires are presented will minimize incorrect responses and/or questionnaire completion time. Thus, researchers may want to limit the reading level of questions presented as part of the assessments (eg, eighth grade reading level). Language fluency is particularly important in studies using samples that are bilingual.

Previous Experience Using Mobile Devices. One potential challenge to recruitment for studies using mobile devices is the participants' potential concerns about their ability to successfully use the assessment device due to lack of knowledge or experience with computers. Therefore, upon initial meetings with potential participants, it is recommended to present the actual device to be used, to describe the types of questions to be presented (ie, "the questions will focus on your mood and social environment"), and/or to allow participants to view a few of the questions on the device's screen. Despite patient concerns, participation and completion rates in studies using mobile devices were comparable to paper-based assessments. For example, Swendsen et al²⁹ reported that of 199 community-dwelling individuals with schizophrenia who were contacted to participate in the study, 92% accepted and 73% completed the study. Among participants who completed the study, the average response rate for the multiple daily life assessments was 72%. This rate is also similar to those found using mobile devices in other samples with schizophrenia of 81% and 79%.^{13,32} Moreover, lack of previous experience using mobile devices does not appear to significantly influence ability to participate in studies using such devices. For example, Granholm et al¹⁴ reported that over 90% of 54 participants in their study of middle-aged individuals with schizophrenia living in the community had no prior experience using electronic mobile monitoring devices. Similarly, hospitalized participants with psychosis displayed few difficulties using mobile devices.^{13,32} Kimhy et al¹³ reported no significant differences between hospitalized participants with psychosis and healthy controls in rating their ability to understand the presented questions, type responses or operate the mobile devices, their stress level, or their level of comfort carrying the mobile devices.

Participants with psychosis also typically express willingness to participate in mobile monitoring studies in the future. However, these individuals may characterize their participation in the study as significantly more challenging than normal controls.¹³ It is therefore important to

note that many previous investigations have included strategies to simplify user interfaces. For example, Granholm et al¹⁴ divided questions with numerous potential responses into multiple questions with fewer response options in order to avoid the need for scrolling within a given screen. Such needs for alterations depend on the nature of the sample being studied and can be readily identified through pilot testing.

Compensation. Depending on the setting and local institutional guidelines, researchers may elect to compensate participants for their participation in mobile assessments. While providing such monetary compensation may potentially increase recruitment and compliance, researchers should limit compensation amounts so as to not introduce coercion (as for any investigation). However, if monetary compensation is offered, the bulk of it should be based on successful completion of the study and on returning the equipment, with a limited portion being based of the compliance.

Issues to Address Prior to Mobile Monitoring

Participant Training in Mobile Monitoring. Whether using paper-based methodologies or mobile devices, researchers need to provide participants with an introductory training session or a tutorial on study procedures. In most published investigations, such sessions occurred on the day of the assessment or the day before, and they were conducted individually over a period of approximately 30 minutes. This introduction may include a description of the assessment schedule and the duration of monitoring, the daily start and end time, the maximal response time to the initial question before the device turns off (if applicable), and the time it typically takes to complete each questionnaire. The introduction session may also allow participants first-hand experience with completing the questionnaire. Such tutorials may reduce participants' stress and potential apprehension about completing the assessment, ensure participants understand all questions, and increase the probability of successful task completion. Importantly, researchers may also want to ask participants not to respond to questionnaires under certain conditions that may lead to potential injuries (eg, while driving, crossing a street, etc). Finally, participants should be provided with contact information to call in the event of potential problems.

Considerations for Scheduling Assessments. In scheduling assessments, researchers may want to take into account the participants' personal daily schedule. While a uniform daily assessment schedule (eg, 10 AM–10 PM) for all participants is methodologically more attractive, a flexible schedule adapted to each participant's unique schedule (for example—1 PM–1 AM for a participant who regularly wakes up at noon and goes to sleep at 3 AM) may offer potentially higher compliance rates and more accurately characterize the true

Table 1. Issues to Address in the Design and Execution of Mobile Assessment Research

Methodological Question	Issues for Researchers to Decide
Sampling strategies	<p>Event vs time sampling Assessments administered only when specific events/experiences occur or based on a predetermined schedule?</p> <p>Random vs fixed time sampling Assessments of experiences at random times or based on a schedule with fixed time points?</p> <p>Number of days and assessments How many days of mobile assessment? How many hours per day could potentially be assessed? How many assessments per day?</p>
Questionnaires development	<p>Constructing questionnaires Are the responses to items expected to vary across time and/or context? Does the language reflect how individuals communicate during daily functioning? Are the items negatively formulated? Do the items require reflection?</p>
Devices and software	<p>Selecting devices and software Is the device used with paper-and-pencil methods or does it administer an electronic interview? Does the device offer potential distractions (ie, games, etc.)? Is the cost of the device prohibitive if lost/damaged? Does operating the device require Information technology support?</p>
Sample recruitment	<p>Clinical considerations Are there any clinical issues that may prevent patients' participation (ie, disorganization, heavy sedation from medications; poor eye sight)?</p> <p>Considerations for inpatient vs outpatient samples For inpatients studies, has unit's clinical staff cooperation been obtained? Do the mobile assessments conflict with clinical appointments?</p>
Sample characteristics and study compliance	<p>Language Are the participants fluent in the questionnaire's language?</p>

Table 1. Continued

Methodological Question	Issues for Researchers to Decide
Compensation	<p>Compensation Does the compensation introduce elements of coercion? Is the bulk of payment more heavily concentrated on return of the mobile device?</p>
Issues to address prior to mobile monitoring	<p>Participant training in mobile monitoring Are participants sufficiently trained to complete the ambulatory monitoring?</p> <p>Considerations for scheduling assessments Schedule the mobile assessments using uniform program or adapt to individual participants? If the mobile assessments include weekends, are they adapted to the participant's schedule?</p>
Issues to address during mobile monitoring	<p>Patient contact How often should the researcher contact participants during the mobile assessment period?</p>

variation observed in the sample. Additionally, if applicable, researchers may want to consider altering the daily assessment schedules for weekends to allow for participants' unique daily schedules during such days.

Issues to Address During Mobile Monitoring

In many previous investigations, the researcher contacted participants by phone or in person (for inpatient studies) during the mobile collection phase, especially at the start of the study, in order to ensure that the first few questionnaires were successfully completed. Such contacts should be very brief to minimize influence on participants' behavior and should be focused on troubleshooting (if necessary), reinforcing compliance, and ensuring that the participants are progressing smoothly in the assessments. Depending on the length of the monitoring period, additional contacts may be advised. Table 1 presents a summary of the diverse issues to be considered in the design and execution of mobile assessments.

Issues to Address After Mobile Monitoring

Researchers may also want to conduct debriefing after the completion of the mobile assessment—such debriefing may allow participants to report problems they may have encountered, as well as to report potential data entry errors (eg, “in the afternoon of the first day, I mistakenly

indicated that I used marijuana”). This period is also ideal to collect information on overall experiences, positive or negative, with the mobile methodology and information about the reasons for eventual compliance difficulties.

Statistical Issues and ESM/EMA Data

Despite its multiple advantages, the structure of repeated measures data complicates its statistical analysis. In this section, we provide a general introduction to data management strategies and statistical approaches commonly used in this area of research.

Initial Considerations

It is likely that mobile assessment studies will include a large amount of missing data. Typically, only participants completing a third or more assessments over the sampling procedure are included in the final analyses,¹⁶ but more lenient inclusion criterion have also been set.³⁴ It may be useful to conduct a sensitivity analysis in order to evaluate whether the inclusion of participants with less entries significantly alters the findings. Researchers should examine whether the main sociodemographic or clinical characteristics of the sample significantly predict the number of completed entries, which could help detect sample biases and aid the interpretation of the results.²⁹

Momentary scales are often skewed, which means that the normality assumption is not always met. In the presence of a conceptually meaningful cutoff point (eg, the presence or absence of phenomena), it may therefore be appropriate to transform continuous variables into binary form.³⁵ Alternatively, this can be accounted for by using robust SEs or bootstrapping in order to introduce conservative CIs.³⁶ It should be noted, however, that this latter option can substantially lengthen the time required to analyze the data.

While most statistical models examine mean differences, researchers may also be interested in the variability or instability of phenomena. For example, one study found the instability of mood over time to be a better predictor of suicidality in individuals with psychosis than was the overall mood level itself.³⁴ The most commonly used metric of instability is the SD,³⁰ although this fails to distinguish between a truly fluctuating course and a single shift from low to high or high to low. The mean squared successive difference (MSSD) provides a more precise measure of instability, but it is influenced by the length of time between assessments.³⁷ In practice, SD and MSSD scores often provide similar findings (J. Palmier-Claus, N. Shryane, P. Taylor, S. Lewis, and R. Drake, unpublished data). Metrics commonly used in experimental research (eg, reciprocals of scores, skewness scores) may also be applicable to the analysis of such data but require large numbers of observations for each participant.³⁸

Examining the instability of variables provides little insight into the direction of change or temporal relationships between phenomena. In order to assess these associations, researchers may create new variables representing the individual's score at the previous time point. In most forms of analyses, this can be achieved by simply moving a column up or down a cell in the dataset and deleting data for the first or last entry of each day when in a long format so that the analyses do not mix within and across-day associations. Such “time-lagging” of the data provides essential information regarding the temporal patterns underlying observed associations and therefore constitutes a prerequisite for testing hypotheses of causality. However, even when adjusting for relevant covariates at multiple time points in the repeated measures data, the presence of time-dependent confounding could invalidate the use of many statistical methods. Time-dependent confounding occurs when a covariate predicts future exposure and future outcomes conditional on past exposure.³⁹

Statistical Approaches to Analyzing Repeated Measures Data

Previous publications^{17,40} have discussed the nested structure of mobile assessment data, which violates the assumption of independent observations (eg, standard approaches to regression analysis, analysis of variance). Independence assumes that observations are uncorrelated with one another, but this is often not the case when multiple observations are generated by the same individual. In such data one might expect nesting to occur at 3 levels: (1) within participants, (2) within days, and (3) at the entry number. For example, at the second level, mood scores might be more strongly correlated within any given day than between days. Averaging the multiple scores for each participant overcomes this problem but loses all information about within-person variation and may be a less sensitive measure of change over time. For these reasons, many researchers now use a technique called multilevel modeling (MLM) to analyze mobile assessment data. MLM is advantageous in that it estimates the level of variance that occurs within and between the levels of the analysis. For example, it provides estimates of the amount of variance that can be accounted for within an individual over time but also that which exists between different individuals. In the case of momentary data, it allows the researcher to observe the effects that the assessment day, the participant, or other variables are having on the dependent variable. The interpretation of the fixed part of the model is then similar to that of standard linear regression or logistic regression (binary outcomes). Although previous studies have explicitly accounted for each level of nesting in data generated by repeated mobile assessments (assessment number, day of the study, and participant ID number), incorporating the highest level in which data are nested (ie, participant ID) should provide equivalent outcomes

for the fixed-effects part of the model. MLM is also appropriate in that it can use maximum likelihood estimations, which means that all measured observations will be included even if missing at some data points. However, depending on the study design, researchers may wish to model the time of entries rather than responses themselves. This approach allows for more accurate estimations of the temporal association between variables, especially when random or semisampling sampling procedures are employed.

Concerning other options in the statistical analysis of mobile assessment data, regression with clustering often provides similar estimates to the fixed-effects part of MLM analysis.⁴¹ A subtype of MLM, Latent Growth Curve Modeling also allows for the impact of a variable on the development or growth of phenomena over time⁴² and may therefore be useful in examining the factors influencing change (eg, treatment effects). Survival analysis may also be appropriate to estimate the proportion of individuals likely to endorse an event at some point in the sampling protocol.⁴³ For example, it could be used to estimate the probability of relapse as well as the factors influencing the chances of its occurrence.

Power Calculation

When designing a study, a sample size or power calculation is often needed to determine the number of participants that are required to confirm or reject a hypothesis. A power calculation allows us to determine the amount of data required in order to detect whether or not a relationship (effect) exists in a population. Such calculations are complicated in MLM because statistical power is determined at more than one level. For example, multilevel power raises the question of whether it is appropriate to measure relatively few time points over many participants or vice versa, as well as many other decisions that affect research design. Essentially, this should be determined according to the level at which the variable of interest is located. If the researcher is interested in the relationship of a variable across time points, the power is determined by the number of assessment entries. However, enough participants should be included to ensure results can be generalized. Alternatively, if associations across days are of primary interest, this should determine the overall sample size.⁴⁴ Commonly used programs for multilevel power calculations include PinT,⁴⁵ Optimal Design,⁴⁶ and RMASS2.⁴⁷ As the pertinence of each of these programs varies according to individual study design and research questions, an increasingly used and flexible option is that of simulations in Mplus as described by Muthén and Muthén.⁴⁸

Conclusions

The feasibility and validity of mobile assessments relative to traditional hospital-based measures have been widely

demonstrated, and previous investigations using this approach in psychiatry have been able to examine a diverse range of clinical issues that are often inaccessible to standard clinical protocols. However, the application of mobile assessment in schizophrenia is particularly complex and requires careful clinical and technical preparation. This article reviews the key decisions for researchers using this method including study length, the nature and timing of assessments, electronic interview content, and device selection. In light of the severe nature of schizophrenia, it also addresses major approaches to sample selection, recruitment, and retention strategies that have been used with success in this population. Finally, several data management and statistical issues are discussed for analyzing complex datasets that reflect both within and between-person variation. The growing use of mobile technologies worldwide will likely place increasing importance on their applications in health care research, both as a research tool and as a means of providing interventions at the most useful moments of the patient's daily life.

Acknowledgments

The authors have declared that there are no conflicts of interest in relation to the subject of this study.

References

1. National Institutes of Health. *Mobile Technologies and Health Care*, Vol 5. Bethesda, MD: NIH Medline Plus; 2011:2–3.
2. Larson R, Csikszentmihalyi M. The experience sampling method. *New Dir Methodol Soc Behav Sci*. 1983;15:41–56.
3. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol*. 2008;4:1–32.
4. Delespaul P, deVries M. The daily life of ambulatory chronic mental patients. *J Nerv Ment Dis*. 1987;175:537–544.
5. Barge-Schaapveld DQ, Nicolson NA, van der Hoop RG, De Vries MW. Changes in daily life experience associated with clinical improvement in depression. *J Affect Disord*. 1995;17:139–154.
6. Loewenstein R, Hamilton J, Alagna S, Reid N, deVries M. Experience sampling in the study of multiple personality disorder. *Am J Psychiatry*. 1987;144:19–24.
7. Myin-Germeys I, Krabbendam L, Jolles J, Delespaul PA, van Os J. Are cognitive impairments associated with sensitivity to stress in schizophrenia? An experience sampling study. *Am J Psychiatry*. 2002;159:443–449.
8. Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, van Os J. Experience sampling research in psychopathology: opening the black box of daily life. *Psychol Med*. 2009;12:1–15.
9. Myin-Germeys I, Peeters F, Havermans R, et al. Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study. *Acta Psychiatr Scand*. 2003;107:124–131.
10. Swendsen J. Anxiety, depression, and their comorbidity: an experience sampling test of the helplessness-hopelessness theory. *Cogn Ther Res*. 1997;21:97–114.

11. Swendsen J. The experience of anxious and depressed moods in daily life: an idiographic and cross-situational test of the Helplessness-Hopelessness Theory. *J Pers Soc Psychol.* 1998; 74:1398–1408.
12. Verdoux H, Gindre C, Sorbara F, Tournier M, Swendsen J. Effects of cannabis and psychosis vulnerability in daily life: An experience sampling test study. *Psychol Med.* 2003;33:23–32.
13. Kimhy D, Delespaul P, Corcoran C, Ahn H, Yale S, Malaspina D. Computerized experience sampling method (ESM): assessing feasibility and validity among individuals with schizophrenia. *J Psychiatr Res.* 2006;40:221–230.
14. Granholm E, Loh C, Swendsen J. Feasibility and validity of computerized ecological momentary assessment in schizophrenia. *Schizophr Bull.* 2008;34:507–514.
15. Delespaul PAEG. *Assessing Schizophrenia in Daily Life.* Maastricht, The Netherlands: Universitaire Pers Maastricht; 1995.
16. Myin-Germeys I, van Os J, Schwartz JE, Stone AA, Delespaul PA. Emotional reactivity to daily life stress in psychosis. *Arch Gen Psychiatry.* 2001;58:1137–1144.
17. Palmier-Claus JE, Myin-Germeys I, Barkus E, et al. Experience sampling research in individuals with mental illness: reflections and guidance. *Acta Psychiatr Scand.* 2011;123: 12–20. <http://mhens.unimaas.nl/index.php?ID=1124>.
18. Myin-Germeys I. Psychiatry. In: Mehl MR, Conner TS, eds. *Handbook of Research Methods for Studying Daily Life.* New York, NY: Guilford Press; 2011.
19. Wichers M, Hartmann JA, Kramer IM, et al. Translating assessments of the film of daily life into person-tailored feedback interventions in depression. *Acta Psychiatr Scand.* 2011;123:402–403.
20. Myin-Germeys I, Birchwood M, Kwapil T. From environment to therapy in psychosis: a real-world momentary assessment approach. *Schizophr Bull.* 2011;37:244–247. <http://www.psymate.eu/>. Accessed October 15, 2011.
21. Johnson EI, Grondin O, Barrault M, et al. Computerized ambulatory monitoring in psychiatry: a multi-site collaborative study of acceptability, compliance, and reactivity. *Int J Methods Psychiatr Res.* 2009;18:48–57.
22. Green AS, Rafaeli E, Bolger N, Shrout PE, Reis HT. Paper or plastic? Data equivalence in paper and electronic diaries. *Psychol Methods.* 2006;11:87–105.
23. Weiss HM, Beal DJ, Lucy SL, MacDermid SM. Constructing EMA studies with PMAT: The Purdue Momentary Assessment Tool User's Manual. Military Family Research Institute, Purdue University; 2004. <http://www.ruf.rice.edu/~dbeal/pmatusermanual.pdf>. Accessed October 1, 2011.
24. My Experience Open Source Software Under the BSD License. 2011; <http://myexperience.sourceforge.net/>. Accessed June 22, 2011.
25. Lewis. Personal Correspondence. 2011. <http://myexperience.sourceforge.net>. Accessed June 22, 2011.
26. Delespaul P, deVries M, van Os J. Determinants of occurrence and recovery from hallucinations in daily life. *Soc Psychiatry Psychiatr Epidemiol.* 2002;37:97–104.
27. Henquet C, van Os J, Kuepper R, et al. Psychosis reactivity to cannabis use in daily life: an experience sampling study. *Br J Psychiatry.* 2010;196:447–453.
28. Myin-Germeys I, Nicolson NA, Delespaul PA. The context of delusional experiences in the daily life of patients with schizophrenia. *Psychol Med.* 2001;31:489–498.
29. Swendsen J, Ben-Zeev D, Granholm E. Real-time electronic ambulatory monitoring of substance use and symptom expression in schizophrenia. *Am J Psychiatry.* 2011;168:202–209.
30. Thewissen V, Bentall RP, Lecomte T, van Os J, Myin-Germeys I. Fluctuations in self-esteem and paranoia in the context of daily life. *J Abnorm Psychol.* 2008;117:143–153.
31. Kimhy D, Corcoran C. Use of Palm computer as an adjunct to cognitive-behavioural therapy with an ultra-high-risk patient: a case report. *Early Interv Psychiatry.* 2008;2:234–241.
32. Kimhy D, Delespaul P, Ahn H, et al. Concurrent measurement of “real-world” stress and arousal in individuals with psychosis: assessing the feasibility and validity of a novel methodology. *Schizophr Bull.* 2010;36:1131–1139.
33. Depp CA, Mautsach B, Granholm E, et al. Mobile interventions for severe mental illness design and preliminary data from three approaches. *J Nerv Ment Dis.* 2010;198: 715–721.
34. Palmier Claus J, Taylor P, Gooding P, Dunn G, Lewis S. Affective variability predicts suicidal ideation in individuals at ultra high risk of developing psychosis: an experience sampling study. *Br J Clin Psychol.* 2011. In press.
35. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ.* 2006;332:1080.
36. Mooney CZ, Duval RD, Duval R. *Bootstrapping: A Nonparametric Approach to Statistical Inference.* Newbury Park, CA: Sage Publications, Inc; 1993.
37. Ebner-Priemer UW, Eid M, Kleindienst N, Stabenow S, Trull TJ. Analytic strategies for understanding affective (in) stability and other dynamic processes in psychopathology. *J Abnorm Psychol.* 2009;118:195–202.
38. Saville CWN, Pawling R, Trullinger M, et al. On the stability of instability: optimising the reliability of intra-subject variability of reaction times. *Pers Ind Differ.* 2011;51:148–153.
39. Hernán MÁ, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology.* 2000;11:561.
40. Palmier Claus J, Myin Germeys I, Barkus E, et al. Clinical overview: experience sampling research in individuals with mental illness: reflections and guidance. *Acta Psychiatr Scand.* 2010;123:12–20.
41. Schwartz JE, Stone AA. The analysis of real-time momentary data: a practical guide. In: Stone AA, Shiffman S, Atienza AA, Nebeling L, eds. *The Science of Real-Time Data Capture: Self-Reports in Health Research.* New York, NY: Oxford University Press; 2007;76–113.
42. Rogers WH. Regression standard errors in clustered samples. *Stata Tech Bull.* 1993;13:19–23.
43. Singer JD, Willet JB. A framework for investigating change over time. In: Singer JD, Willet JB, eds. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* New York, NY: Oxford University Press; 2003;3–15.
44. Allison PD. *Survival Analysis Using SAS: A Practical Guide.* Cary, NC: SAS Institute; 2010.
45. Snijders TAB. Power and sample size in multilevel modeling. *Encyclopedia stat behav sci.* 2005;3:1570–1573.
46. Raudenbush SW, Xiao-Feng L. Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychol Methods.* 2001;6:387–401.
47. Hedeker D, Gibbons RD, Watnaux C. Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. *J Educ Behav Stat.* 1999;24:70–93.
48. Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. *Struct Equation Model.* 2002;4:599–620.