Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology

Debbie S. Moskowitz, PhD; Simon N. Young, PhD

Moskowitz — Department of Psychology; Young — Department of Psychiatry, McGill University, Montréal, Que.

Current methods of assessment in clinical psychopharmacology have several serious disadvantages, particularly for the study of social functioning. We aimed to review the strengths and weaknesses of current methods used in clinical psychopharmacology and to compare them with a group of methods, developed by personality/social psychologists, termed ecological momentary assessment (EMA), which permit the research participant to report on symptoms, affect and behaviour close in time to experience and which sample many events or time periods. EMA has a number of advantages over more traditional methods for the assessment of patients in clinical psychopharmacological studies. It can both complement and, in part, replace existing methods. EMA methods will permit more sensitive assessments and will enable more wide-ranging and detailed measurements of mood and behaviour. These types of methods should be adopted more widely by clinical psychopharmacology researchers.

Les méthodes actuelles d'évaluation en psychopharmacologie clinique présentent plusieurs inconvénients sérieux, en particulier pour l'étude du fonctionnement social. Nous voulions étudier les forces et les faiblesses des méthodes actuellement utilisées en psychopharmacologie clinique et les comparer à un ensemble de méthodes mises au point par des psychologues spécialistes de la personnalité et des sociopsychologues, appelé évaluation écologique ponctuelle (EEP), et qui permet au participant à la recherche de signaler symptômes, affect et comportements presque au moment où il les vit et d'échantillonner de nombreux événements ou périodes. L'EEP offre de nombreux avantages sur les méthodes plus traditionnelles d'évaluation des patients au cours d'études en psychopharmacologie clinique. Elle peut à la fois compléter des méthodes existantes et les remplacer en partie. Les méthodes d'EEP permettront d'effectuer des évaluations plus sensibles et des mesures de plus grande envergure et plus détaillées d'humeur et de comportement. Les chercheurs en psychopharmacologie clinique devraient adopter ces méthodes en plus grands nombres.

Introduction

The purpose of the present commentary is to consider the limitations of the primary measurement strategies used to study affect, mood and interpersonal behaviour in psychopharmacological studies. We propose that a set of methods derived from personality/social psychology research, and collectively referred to as ecological momentary assessment (EMA), has several important advantages over currently used methods. EMA techniques provide methods by which a research participant can report on symptoms, affect, behaviour and cognitions close in time to experience, and these reports are obtained many times over the course of a study. We argue that these methods should be used more widely in studies of treatment effectiveness. EMA methods should improve the measurement of many of the common outcomes of psychopharmacological studies, such as mood and anxiety. They also permit the study of human social interaction in a way that is not possible with the current methodology. This commentary focuses primarily on the potential of EMA for the measurement of social interactions.

Disturbances of social functioning are an important component of many types of psychopathology. Different types of psychopathology alter social interactions in different ways,

Correspondence to: Dr. Debbie S. Moskowitz, Department of Psychology, McGill University, Stewart Biology Bldg., 1205 Dr. Penfield Ave., Montréal QC H3A 1B1; fax 514 398-4896; dsm@ego.psych.mcgill.ca

Medical subject headings: affect; clinical trials; interpersonal relations; psychology, social; research.

J Psychiatry Neurosci 2006;31(1):13-20.

Submitted Nov. 1, 2004; Revised June 11, 2005; Accepted Sept. 6, 2005

and symptoms that impair an individual's ability to have social interactions can impair that individual's subjective well-being and quality of life. Individuals with psychopathology may misperceive the intent of others; they may behave inappropriately, and they may have unusual affective responses in their interactions with others. For example, individuals who have high scores on neuroticism, a characteristic found in a variety of psychiatric disorders such as the personality disorders,[1] have been found to exhibit high rates of both submissive and quarrelsome behaviours. They also have unusual affective responses to behaving in a submissive and quarrelsome manner.[2] Depression and anxiety are the most common forms of psychopathology in North America. Both conditions can have symptoms that manifest themselves during the course of social interactions. Individuals who are depressed talk less; they make more hostile comments and fewer positive comments, and they have problems with excessive submission.[3] Socially anxious individuals fear social or performance situations; social interactions are experienced with intense subjective distress and are frequently avoided. These individuals have problems engaging in a smooth flow of social interaction with others.[4,5] Thus, when considering measurement strategies for clinical psychopharmacology that can elucidate the full range of symptoms and manifestations of a disorder, it is important to consider measurement of aspects of interpersonal interactions as well as mood and affect. This commentary is focused on the measurement of social interaction, because the typical measurement of social interaction is impoverished and the implementation of EMA methods has the potential to provide an understanding of social interaction that has eluded present measures.

Current measurement of social behaviour

To determine what is currently being measured in clinical psychopharmacology studies, we surveyed the 8 journals with the highest impact factors in the psychiatry listing in the 2003 ISI Journal Citation Reports. We examined clinical psychopharmacological studies published during 2004 to determine the kinds of methods used to measure mood and affect and social behaviour. The details are presented in Table 1. For each study involving long-term treatment with a psychopharmacological agent, the following information was recorded: whether the study involved the assessment of social behaviour, and for purposes of contrast also mood and affect, and the kind of method (e.g., clinician rating, self-report, report by a knowledgeable informant, laboratory test, diary or EMA-type method) used to assess these domains. Case studies, chart reviews, reports of secondary analyses and reviews involving meta-analyses were not included.

The primary method reported consisted of clinicians' judgements based on observation of the subject or information provided by the subject to the clinician. A clinician report was used more than twice as often as the next most commonly used method, questionnaires completed by the research participant. A few studies included other methods: reports by knowledgeable informants such as caregivers, spouses, friends and teachers; tests in the laboratory including observations by trained raters under standardized conditions; and diary or EMA-type methods.

Almost twice as many studies examined mood and affect compared with the number of studies examining social behaviour. Studies of mood and affect usually used instruments that assessed multiple aspects of mood and affect. In contrast, studies of social behaviour frequently involved only a global assessment of social functioning, without detailed assessment of aspects of social functioning, qualities of social interactions in which the person was involved, characteristic behaviours of the person in social interactions, or characteristics of the other person or the situations that might have an impact on a person's symptoms. Despite the centrality of social behaviour to descriptions of psychopathology, social behaviour is examined less frequently and in less detail in psychopharmacological studies than mood and affect.

How effective are current methods in studying these content domains? In considering the usefulness of methods, it is important to consider common threats to their validity.[6–8] One important threat to validity is a poor fit between the method and the definition of an outcome variable. Examples of poor fits would include the use of a measure with low test–retest reliability to assess an outcome theoretically defined to be stable over occasions, the use of a measure consisting of items that refer to 2 or 3 features of an outcome variable to assess a characteristic with a large number of features, and the use of a measure that is specific to one situation to assess an outcome that is expected to have generality across a variety of situa-

Table 1: Methods used to measure mood/affect and social behaviour in psychopharmacological studies*

	Method used						
Content domain	Clinician rating	Self-report	Other report	Laboratory test	EMA	Other	Total no. of studies
Mood/affect	163	61	4	2	4	1	177
Social behaviour	85	19	5	3	2	1	97
Total no. of studies using method	164	64	6	5	4	1	180

EMA = ecological momentary assessment.
*The journals surveyed were *Archives of General Psychiatry*, *American Journal of Psychiatry*, *Biological Psychiatry*, *Molecular Psychiatry*, *Neuropsychopharmacology*, *Journal of Clinical Psychiatry*, *Journal of Clinical Psychopharmacology* and *British Journal of Psychiatry*. A total of 207 studies were identified that manipulated a psychopharmacological agent. Of these studies, 27 did not measure mood/affect or social behaviour. The 180 studies that did examine these content domains often examined more than one content domain and used more than one method.

tions. For example, a patient who has moods that tend to fluctuate might focus on recent mood in response to a clinician's inquiry, "How have you been feeling," and not give an accurate estimate of mood over the period for which the query was intended to solicit information. Inquiring about a subject's overall social adaptation may not elicit information about the person's range of social adaptation at work, at home, with friends and in necessary real-life encounters such as with doctors and administrative clerks. Additional threats to validity can stem from the respondent and from the person collecting the information. There can be threats stemming from the respondent if the person is reactive to the measurement procedure or the respondent is unable to complete the data-gathering procedure correctly. There may be threats stemming from the data gatherer, if the data gatherer is unable to complete the data collection procedure accurately or if there are experimenter biases, interviewer effects and changes in the research procedure over time. For example, a subject may not be able to correctly complete a questionnaire; a family member may not be able to report behaviour in situations the family member has not observed; or an interviewer may begin an interview with expectations that preclude the solicitation of relevant information to inform rating judgements. We will consider these threats to the validity of clinician reports, self-reports, laboratory tests and EMA measures specifically focusing on the assessment of social interaction.

Clinician reports

As found in our survey of research reports in journals, reports by clinicians are the most commonly used method in clinical psychopharmacology to study social behaviour and affect. With this method, measures are obtained by collecting information from a clinician who has had some amount of personal contact with the individual.

A variety of formats have been used to quantify information from clinicians. Clinicians are frequently asked to provide a global rating of improvement or extremity of psychopathology using instruments such as a Clinical Global Impressions (CGI) Scale. This method may not be sensitive to the variety of features of a disorder and, hence, may not provide a good match between the measure and the outcome variable. Indeed, the CGI may rely on a varying synthesis of many different aspects of the functioning of the patient. To provide a better match between measure and outcome variable, clinicians are also generally asked to provide more specific information using detailed rating forms such as the Hamilton Rating Scale for Depression (HAM-D).[9]

Errors stemming from the clinician are potentially a serious threat to the validity of this method. The accuracy of the clinician's information will be affected by the familiarity of the clinician with the individual. Often there is little contact between the research participant and the clinician; the clinician has not had the opportunity to observe the range of relevant behaviours and moods or to observe the individual in the range of situations that might be relevant. In clinical trials of outpatients, the assessment visit will sometimes be the only contact between patient and rater. The clinician is therefore dependent on information provided by the research participant. The limitations of clinicians' assessments are most clearly illustrated by Rosenhan's famous (infamous?) study in which clinicians were unable to detect pseudopatients admitted to a psychiatric ward even though the other patients, with whom they spent much more time, were able to detect their "sanity."[10]

An additional problem with clinicians' ratings is that there often is no standardized procedure to elicit information from the subject, so that judgements are based on different kinds of information across participants. Furthermore, whereas multicentre trials sponsored by drug companies usually include assessments of interrater reliability, sometimes little effort is made to establish reliability across clinicians in other studies. When interrater reliability is established, the most common method is to have different raters view the same taped interview. This practice may examine the extent of consistency in clinicians' perceptions of different people, but it does not examine the consistency with which information is elicited from research participants by different clinicians. In many clinician-rating studies, the extent of agreement among clinicians in the use of measures is not explicitly examined.

There is an additional potential problem in using clinician reports. Most individuals, including clinicians, generally have assumptions about the covariation among different items referring to a characteristic.[11] An implication of these implicit beliefs can be that if the clinician rates change on one item, the clinician may also rate change on other items even if the clinician does not have direct information that the other items have changed. For example, teachers are often used as experts in the evaluation of children's problem behaviours. Schachar et al[12] found that defiant behaviour exerts a halo effect on teachers' evaluations of hyperactivity and inattentiveness such that defiance toward a teacher increases the likelihood that a child will be rated as hyperactive or inattentive, regardless of the level of activity or attentiveness in observed behaviour. Similarly, other expert raters may overemphasize a single symptom in a decision about degree of psychopathology.

The strength of a report by the clinician is that the clinician can combine information elicited from the subject with direct observations of the subject. Unfortunately, the accuracy with which clinicians can combine information has often been found to be far less than would be expected or desired.[13]

Self-report questionnaires

Self-reports are measures derived from the introspection of individuals collected in the form of global ratings or responses to items on questionnaires. Although this method depends on the motivation and ability of the person to answer questions accurately, it can provide an excellent fit with the characteristic of interest. Items can be constructed to assess the full range of behaviours, feelings, sensations and activities reflecting the characteristic under study. When necessary, questions can request information about multiple pertinent situations such as situations involving work, family and other personal relationships (e.g., the Social Adjustment Scale Self-Report[14]).

Systematic influences stemming from the respondent are potentially major threats to the validity of self-report questionnaires and interviews. Many modern questionnaires are constructed with an awareness of the influence of response sets and response biases such as the tendency to acquiesce or the general tendency to respond in a socially desirable way. However, not all response biases can be easily controlled. People reconstruct their recalled memories.[15] Individuals who are highly neurotic, as is the case for many forms of psychopathology, have a greater tendency to recall symptoms.[16,17] The tendency to recall events as more negative than they were experienced at the time is likely to make self-report questionnaires less sensitive to change. For example, to be found to be effective in reducing negative mood and to be in fact reducing negative mood, a psychopharmacological agent would not only have to change the experience of mood on specific occasions but would also have to change the memory of mood upon retrospection.

There are daily and weekly effects on reports of mood and social behaviour.[18–20] Assessments obtained only on particular days or at particular times may be influenced by the timing of the assessment. For example, a depressed patient who shows marked diurnal variation of mood would tend to report differently on overall mood depending on the time of day the questionnaire was completed.

Laboratory tests

With this method, individuals' responses to situations constructed for a laboratory setting are recorded either by trained observers, as in coding of behaviour, or by the individual, as in the reporting of mood. As typically implemented, laboratory measures are based on responses to 1 or possibly 2 brief situations. It is common for researchers to take steps to minimize error stemming from the investigator. For example, situations are standardized, and observers can be trained to a high level of reliability. Reactivity from the respondent is sometimes an issue, but proper counterbalancing of respondents' medication conditions minimizes the influence of changes in reactivity over time on tests comparing 2 psychopharmacological conditions.

In laboratory tests, errors stemming from the match between the method and the definition of the characteristic can be a severe threat to validity. If there is the expectation that the characteristic is limited in its display to certain kinds of situations, then there can potentially be a good fit between the measures and the characteristic by careful selection of the situation or situations. However, laboratory situations are often artificial in the sense that the stimuli are unlike any sets of stimuli that would ordinarily be encountered in everyday life (i.e., limited generalizability). Moreover, the kinds of responses permissible may not be like the responses of ordinary life or may not adequately reflect the full range of responses to everyday or more extreme events. Each laboratory situation is limited in the range of behaviour likely to be evoked. For example, when studying cooperativeness, a researcher might be able to create a task that would permit 1 or 2 signs of cooperativeness, such as sharing materials, taking turns, preparing joint plans or maximizing the group's rather than the individual's gains, but it is unlikely that 1 situation could be created that would permit an individual to reasonably engage in all these behaviours. So the use of a laboratory situation permits the assessment of a few judiciously selected behaviours to reflect a characteristic, but the behaviours observed are not likely to match the range of behaviours subsumed by the characteristic.

A commonly used laboratory task of impulsivity is the Go/No Go task, in which subjects have to respond to some stimuli by pressing a button and withhold their response to other stimuli. The measure of impulsivity is the number of commission errors, that is, pressing the button when the subject should not. Increased commission errors are found in patients with disorders associated with lack of impulse control.[21–23] However, agreement between the Go/No Go task and other behavioural and paper and pencil tests of impulsivity across subjects is weak.[24] Thus, the relation between a laboratory test such as the Go/No Go task and real-life situations in which impulsive behaviour has an adverse impact on an individual, such as a sudden expression of anger toward another person, is not clear.

Furthermore, the temporal reliability (reliability over occasions) of laboratory measures is a concern.[25,26] If the measure is not reliable over occasions, then the measure is associated with considerable error variance that will reduce the sensitivity of the measure to change that might be induced by a psychopharmacological agent.

An alternative: from daily sampling to EMA

As previously noted, methods based on self-reports have many advantages in being able to measure a range of moods and affects and a range of interpersonal behaviours in a variety of situations. The major disadvantage is that recalled information is influenced by reconstructive processes that reduce its accuracy. Many contemporary methodologists argue that collecting self-reported information closer in time to its occurrence will reduce the reliance on memory and consequently improve accuracy. Several approaches have been taken to collecting measures close in time to experience that are examples of EMA.[27–29]

One approach has been the collection of information at specified time intervals, such as once-a-day diaries. These measures are typically completed at the end of the day and ask individuals to look back on the events of the day in reporting about variables such as mood, stress and stressors, and relations with others. To reduce the reliance on memory even further, other methods require the reporting of information several times a day. One approach to reducing the time between an experience and the report of the experience is signal-contingent recording, in which reports are requested in response to a signal that occurs a fixed number of times per day on a random schedule. A second approach is event-contingent recording, in which reports are requested in response to specifically identified events, such as interpersonal interactions. A strength of signal-contingent recording is that all participants are given the same number of signals and hence

report upon (with the exception of missing data) the same number of events. A weakness of signal-contingent recording is that events important to the researcher may be missed. For example, if someone is socially anxious and has only a few social interactions in a day, all these events may be missed if signals do not co-occur with the times of the social interactions. Important events can be defined to instigate their recording.

This kind of methodology has been used with samples from many populations, including healthy working adults, patients with chronic pain, depressed individuals and persons vulnerable to psychosis.[30–35] Variations of this methodology have been found to be suitable for use with individuals from a broad age range, including children, adolescents and the elderly.[36–38] There have been many demonstrations of the reliability and validity of measures based on EMA methodologies. Research findings have supported the reliability of measures aggregated across occasions (temporal stability) and the reliability across items within a measure (internal consistency) for measures of positive and negative valence of affect and measures of specific kinds of affect such as joy, fear and sadness.[39,40] Reports of pain using this kind of method have been found to correlate with assessments of pain sensitivity in the laboratory.[41] The data are sensitive to change in drugs or medication.[42,43]

Several approaches have been taken to the use of these data. The data can be aggregated to produce a single score for each condition, psychopharmacological agent or placebo. Most importantly, though, there are multiple data points. Aggregation of multiple data points decreases the error variance in the measure (see Epstein[44] and Moskowitz and Schwarz[45]). Reduction of error will mean that the measures are much more sensitive to systematic change such as could be produced by a drug. Aggregation can be done explicitly as in taking the means across observations, or the multiple data points can be considered as a set as in the case of multilevel and random coefficient modelling.[46–48]

The fact that there are multiple data points collected across time can be further exploited. With numerous time points, it is possible to look at the range of scores for an individual on a variable, for example, to provide a measurement of mood lability or behavioural variability.[49,50] This could be done as a general phenomenon, or in response to specific events, for example, the range of moods in patients with bulimia and temporal change after binges. The slope of change across the data-collection period can be calculated to see if the pattern of change is different for one kind of behaviour or affect than for others. Depending on the length of the time period sampled, this might further permit the examination of the ordering of change, for example, whether change in affect precedes or is subsequent to change in behaviour.

This method lends itself to the examination of other questions concerning time-dependent sequences related to the use of psychopharmacological agents, for example, whether affect, behaviour and environmental events such as daily stressors mediate the need for greater or lesser medication or drug use (e.g., Armeli et al[51]). Moreover, the relations of variables within subjects can be assessed, and this relation may be affected by the pharmacological agent. For example, in depressed patients, the temporal relation between improvements in mood and improvements in different aspects of social interaction (e.g., duration and degree of agreeableness) could be examined. It is known that the relation between affect and interpersonal behaviour is much weaker for individuals with high scores on neuroticism than for individuals with low scores.[2] The relation between behaviour and affect may also be less strong for individuals with some forms of psychopathology than for healthy individuals, and this relation may become stronger with proper treatment.

Is it feasible and practical to use EMA in patients with psychiatric disorders?

EMA procedures are more demanding of research participants than more global types of assessment. Their application in psychopharmacological studies will depend on the willingness and ability of patients to carry out successfully all the procedures that these methods require. Evidence to date suggests that most patients should have no problems doing this. Although the number of psychopharmacological studies in which EMA has been used is very limited, EMA methods have been successfully carried out in patients with a variety of different types of psychopathology. For example, impulsivity does not seem to prevent the use of EMA methods, because they have been applied successfully to patients with borderline personality disorder, bulimia, attention-deficit hyperactivity disorder (ADHD) and violent patients. Patients with borderline personality disorder successfully filled in 50 measures over 10 days.[52] In the ADHD study, the participants logged their behaviours, moods and social contexts twice each hour across two 4-day recording intervals.[53] This was done by patients with mild, moderate and severe disturbances. Patients with bulimia recorded their perceptions of social interactions, concurrent self-perceptions and moods, and eating behaviours after each social interaction for up to 22 days.[54] A review of the use of EMA methods in eating disorders concluded that patients are willing and able to engage in EMA studies, and the method makes it possible to collect data that could not be obtained with other study designs.[55] Finally, 25 violent psychiatric patients successfully responded to 50 signals over 7 days by answering a questionnaire with 20 mood state items.[56]

Depressed patients have decreased concentration, memory and motivation, but this does not prevent them from using EMA methods. For example, patients with major depressive disorder reported on their mood and events in their lives 10 times per day for 6 consecutive days.[57] Depressed adolescents and children were also able to use EMA methods.[58] In 2 studies, EMA methods were used in trials of antidepressant drugs. In the first, 21 patients were randomly allocated to receive treatment for 6 weeks with fluvoxamine or amitriptyline. For 12 days, the patients reported on their thoughts, current activity, physical and social context, and mood in response to 10 daily random signals. In the second, 63 patients were treated with imipramine or placebo for 6 weeks. The patients responded to 10 random signals per day for a total of 15 days during treatment by rating their mood, enjoyment of current activity and physical complaints.[59]

A diagnosis of schizophrenia also does not rule out the use of EMA methods. In one study, 42 patients with schizophrenia responded 10 times per day over 6 days by noting their subjective stress of daily life events and disturbances in daily life, as well as their emotional reactivity, as indicated by changes in both negative and positive affect.[34] In another study, 57 patients also reported 10 times per day over 6 days, giving information about their ongoing hallucinations as well as about their thoughts, mood, current activity and social circumstances.[60]

The studies described previously suggest that EMA methods have wide applicability. Of course, there will be individuals with psychopathology who are unable to comply with EMA procedures. However, there are also limitations to who can participate in interviews and questionnaires and who can respond with accuracy to these data-collection procedures. If someone is having problems with memory, then retrospection across a 1-week or 1-month or several-month period as is used in some interviews that are the basis of clinician reports will be less accurate than retrospection across the brief periods used in EMA procedures.

A sample measure of EMA

A method developed by Moskowitz[30] can be used as a case in point for demonstrating the possibilities of EMA for developing reliable and valid measures of behaviour that are sensitive to change. Moskowitz constructed a set of items to correspond with the interpersonal circumplex model of interpersonal behaviour. These items measured 4 dimensions of interpersonal behaviour, namely, dominance, submissiveness, agreeableness and quarrelsomeness, which can be conceptualized as corresponding to 2 independent axes of social behaviour, one referring to status seeking and the other to agonism–affiliation. The specific behaviours selected to represent each dimension were anticipated to occur during daily life, so extreme forms of behaviours such as physical aggression were omitted. The items have been used in an event-contingent recording procedure in which study participants are asked to complete record forms, including a sampling of the behaviours after each significant social interaction, each day for several weeks. Given these instructions, participants typically complete 6–7 record forms throughout each day of a study.

In a series of studies using the method, Moskowitz and associates[2,30,61] have demonstrated that there is a very low dropout rate from these studies. Inter-item reliability for the items on each measure is high. The measures have also been found to be stable across occasions of measurement.[20,30] This temporal stability is dependent on the number of occasions of measurement. Temporal stability increases up to 12 days or about 80 occasions of measurement, at which point temporal stability asymptotes and does not increase further. There have also been demonstrations that the measures of the behavioural dimensions correlate as theoretically predicted with each other and with one-occasion self-report questionnaire measures of these dimensions.[2,30,62] When focusing on specific events, there is convergence between self-reports of the behaviours and reports by individuals interacting with the subject.[63]

The measures are sensitive to change. The measures reflect predicted differences in response to change in situational variables, such as hierarchical status (e.g., whether the person is in the role of boss, coworker or supervisee).[64] Most importantly the measures have been found to be sensitive to the administration of a psychopharmacological agent, tryptophan, as compared with placebo.[61]

Practical aspects of EMA

EMA methodology originated with paper and pencil methods, either with a diary, which is returned to the investigator after a period that may last a week or more, or with single-page questionnaires that are mailed in daily and the postmark verified. This method has also been used with handheld computer devices such as personal digital assistants (PDAs) that are most easily used by young, technology-savvy adults for relatively short data-gathering periods. With improvements in electronics, including permanent memory, increased battery life and improved visibility of screens, the technique is moving to PDAs that will be usable by a broader variety of populations. The advantages of PDAs are that time-contingent methodologies can use a signal programmed into the PDA; the exact time that data are entered in the PDA can be recorded by the program, which may help prevent subjects from entering data through retrospective recall; and the data can be downloaded directly. When PDA technology does not match data-collection requirements or when the research question calls for only infrequent recording, such as once a day, paper and pencil diaries with postmarks as verification should continue to be adequate.

Although it is possible to adopt measures, such as mean scores and variability scores,[50] that can be analyzed using familiar statistical measures, the statistical analysis of EMA methods can be complex. Fully exploiting the potential of the repeated measures obtained in this measurement strategy requires sophisticated data analytic strategies.[29,46–48]

Limitations of EMA

As with all methods, EMA has its disadvantages. It is more time consuming for the research subject than meeting with a clinician at intervals. Another disadvantage of EMA, as with all self-report measures, is that there is no independent check on the veracity of the data, because all data are collected in the absence of the experimenter. One possible way around this is to seek confirmation from others with whom the subject has frequent contact to give reports that can be used to verify at least a subset of the data (see Tse and Bond[65]).

Conclusions

In summary, measures that can be classified under the rubric of EMA methodology have been shown to be reliable and valid, can show excellent matching between the measure and the theoretical definition of the outcome, reduce reliance on retrospective memory and reduce the need to rely on the integrative judgements of either the clinician or the respondent. Because measures based on this class of methodology collect

many data points, they are likely to contain less random error variance than other common methods and, hence, may be more sensitive to change. The collection of multiple data points also permits new questions about changes that may occur as a consequence of a psychopharmacological agent, such as the shape of change, the ordering of change among variables and changing relations among outcome variables.

Measures based on EMA techniques have the potential for assessing a wide variety of outcome variables. Interpersonal behaviour is studied less often than mood and affect. Even when included in a study, the measurement of social behaviour tends to be diffuse and unspecific. EMA such as the event-contingent method described here has the capacity to define social behaviour variables with both great precision and with the necessary breadth to cover several dimensions of social behaviour in a single study. Measurements of mood can be included along with measures of social interaction to assess both kinds of outcomes and the interplay of mood and social behaviour.

So far, clinical psychopharmacology researchers (or indeed psychotherapy researchers) have not adopted EMA methodologies to any appreciable extent. However, the advantages of doing so are clear. Measures will be more refined and, therefore, more sensitive to changes. Possibly fewer patients will be required to detect differences, making studies easier and cheaper to perform. In addition, because EMA methods facilitate the study of detailed aspects of behaviour, questions that are clinically significant but difficult to answer with current methodologies should be easier to examine. For example, it has been claimed that a selective noradrenaline reuptake inhibitor is more effective than a selective serotonin reuptake inhibitor in improving social adaptation, in spite of equivalent clinical efficacy.[66] This is the type of important clinical issue for which EMA methodology is ideal.

The importance of interdisciplinary research is a dogma that is often put forward without any indication of what disciplines should be interacting and how. The marriage of EMA, pioneered by personality/social psychology researchers, with clinical psychopharmacology is an example of a specific interdisciplinary collaboration that holds great promise.

Acknowledgements: This work was supported by grants from the Social Sciences and Humanities Research Council of Canada (to D.S.M.) and from the Canadian Institutes of Health Research (S.N.Y.). We thank Franceen Lenoff for her assistance with reviewing the studies referred to in Table 1, and we thank Howard Tennen for his comments on a draft of this commentary.

Competing interests: None declared.

Contributors: Both authors contributed substantially to drafting and revising the article, and each gave final approval for the article to be published.

References

1. Widiger TA, Costa PT. Personality and personality disorders. *J Abnorm Psychol* 1994;103:78-91.
2. Côté S, Moskowitz DS. On the dynamic covariation between interpersonal behavior and affect: prediction from neuroticism, extraversion, and agreeableness. *J Pers Soc Psychol* 1998;75:1032-46.
3. Alden LE, Phillips N. An interpersonal analysis of social anxiety and depression. *Cognit Ther Res* 1990;14:499-513.
4. Turner SM, Beidel DC. Some further comments on the measurement of social phobia. *Behav Res Ther* 1988;26:411-3.
5. Rapee RM. Descriptive psychopathology of social phobia. In: Heimberg RG, Liebowitz MR, editors. *Social phobia: diagnosis, assessment, and treatment*. New York: Guilford Press; 1995. p. 41-66.
6. Moskowitz DS. Comparison of self-reports, reports by knowledgeable informants, and behavioral observation data. *J Pers* 1986; 54:294-317.
7. Webb EJ, Campbell DT, Schwartz RD, et al. *Unobtrusive measures*. Thousand Oaks (CA): Sage; 2000.
8. Cook TD, Shadish WR, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin; 2001.
9. Hamilton M. Rating depressive patients. *J Clin Psychiatry* 1960; 41:21-4.
10. Rosenhan DL. On being sane in insane places. *Science* 1973;179: 250-8.
11. Kenny DA, Berman JS. Statistical approaches to the correction of correlational bias. *Psychol Bull* 1980;88:288-95.
12. Schachar R, Sandberg S, Rutter M. Agreement between teachers' ratings and observations of hyperactivity, inattentiveness, and defiance. *J Abnorm Child Psychol* 1986;14:331-45.
13. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science* 1989;243:1668-74.
14. Weissman MM, Bothwell S. Assessment of social adjustment by patient self-report. *Arch Gen Psychiatry* 1976;33:1111-5.
15. Schacter DL. *The seven sins of memory: how the mind forgets and remembers*. Boston: Houghton Mifflin; 2001.
16. Larsen RJ. Neuroticism and selective encoding and recall of symptoms: evidence from a combined concurrent-retrospective study. *J Pers Soc Psychol* 1992;62:480-8.
17. Brown KW, Moskowitz DS. Does unhappiness make you sick? The role of affect and neuroticism in the experience of common physical symptoms. *J Pers Soc Psychol* 1997;72:907-17.
18. Larsen RJ, Kasimatis M. Individual differences in entrainment of mood to the weekly calendar. *J Pers Soc Psychol* 1990;58:164-71.
19. Stone AA, Smyth JM, Pickering T, et al. Daily mood variability: form of diurnal patterns and determinants of diurnal patterns. *J Appl Soc Psychol* 1996;26:1286-305.
20. Brown KW, Moskowitz DS. Dynamic stability of behavior: the rhythms of our interpersonal lives. *J Pers* 1998;66:105-34.
21. Newman JP, Kosson DS. Passive avoidance learning in psychopathic and nonpsychopathic offenders. *J Abnorm Psychol* 1986;95:252-6.
22. Newman JP, Patterson CM, Howland EW, et al. Passive avoidance in psychopaths: the effects of reward. *Pers Individ Dif* 1990;11: 1101-14.
23. Iaboni F, Douglas VI, Baker AG. Effects of reward and response costs on inhibition in ADHD children. *J Abnorm Psychol* 1995;104:232-40.
24. Helmers KF, Young SN, Pihl RO. Assessment of impulsivity in healthy male volunteers. *Pers Individ Dif* 1995;19:927-35.
25. Epstein S. The stability of behavior: II. Implications for psychological research. *Am Psychol* 1980;35:790-806.
26. Moskowitz DS. Convergence of self-reports and independent observers: dominance and friendliness. *J Pers Soc Psychol* 1990; 58:1096-106.

27. Wheeler L, Reis HT. Self-recording of everyday life events: origins, types, and uses. *J Pers* 1991;59:339-54.

28. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. *Ann Behav Med* 1994;16:199-202.

29. Affleck G, Zautra A, Tennen H, et al. Multilevel daily process designs for consulting and clinical psychology: a preface for the perplexed. *J Consult Clin Psychol* 1999;67:746-54.

30. Moskowitz DS. Cross-situational generality and the interpersonal circumplex. *J Pers Soc Psychol* 1994;66:921-33.

31. Affleck G, Tennen H, Keefe FJ, et al. Everyday life with osteoarthritis or rheumatoid arthritis: independent effects of disease and gender on daily pain, mood, and coping. *Pain* 1999;83:601-9.

32. Barge-Schaapveld DQ, Nicolson NA, Berkhof J, et al. Quality of life in depression: daily life determinants and variability. *Psychiatry Res* 1999;88:173-89.

33. Schanberg LE, Sandstrom MJ, Starr K, et al. The relationship of daily mood and stressful events to symptoms in juvenile rheumatic disease. *Arthritis Care Res* 2000;13:33-41.

34. Myin-Germeys I, van Os J, Schwartz JE, et al. Emotional reactivity to daily life stress in psychosis. *Arch Gen Psychiatry* 2001;58:1137-44.

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

36. Larson R. Beeping children and adolescents: a method for studying time use and daily experience. *J Youth Adolesc* 1989;18:511-30.

37. Samo JA, Tucker JA, Vuchinich RE. Agreement between self-monitoring, recall, and collateral observation measures of alcohol consumption in older adults. *Behav Assess* 1989;11:391-409.

38. Beidel DC, Neal AM, Lederer AS. The feasibility and validity of a daily diary for the assessment of anxiety in children. *Behav Ther* 1991;22:505-17.

39. Diener E, Emmons RA. The independence of positive and negative affect. *J Pers Soc Psychol* 1984;47:1105-17.

40. Diener E, Smith H, Fujita F. The personality structure of affect. *J Pers Soc Psychol* 1995;69:130-41.

41. D'Antono B, Ditto B, Rios N, et al. Risk for hypertension and diminished pain sensitivity in women: autonomic and daily correlates. *Int J Psychophysiol* 1999;31:175-87.

42. Budney AJ, Hughes JR, Moore BA, et al. Marijuana abstinence effects in marijuana smokers maintained in their home environment. *Arch Gen Psychiatry* 2001;58:917-24.

43. Eardley I, Morgan R, Dinsmore W, et al. Efficacy and safety of sildenafil citrate in the treatment of men with mild to moderate erectile dysfunction. *Br J Psychiatry* 2001;178:325-30.

44. Epstein S. The stability of behavior: I. On predicting most of the people much of the time. *J Pers Soc Psychol* 1979;37:1097-126.

45. Moskowitz DS, Schwarz JC. Validity comparison of behavior counts and ratings by knowledgeable informants. *J Pers Soc Psychol* 1982;42:518-28.

46. Kreft I, De Leeuw J. *Introducing multilevel modelling*. London: Sage; 1998.

47. Kenny DA, Bolger N, Kashy DA. Traditional methods for estimating multilevel models. In: Moskowitz DS, Hershberger SL, editors. *Modeling intraindividual variability with repeated measures data: methods and applications*. Mahwah (NJ): Lawrence Erlbaum Associates; 2002. p. 1-24.

48. Singer JD. Fitting individual growth models using SAS PROC MIXED. In: Moskowitz DS, Hershberger SL, editors. *Modeling intraindividual variability with repeated measures data: methods and applications*. Mahwah (NJ): Lawrence Erlbaum Associates; 2002. p. 135-70.

49. Eid M, Diener E. Intraindividual variability in affect: reliability, validity, and personality correlates. *J Pers Soc Psychol* 1999;76:662-76.

50. Moskowitz DS, Zuroff DC. Flux, pulse, and spin: dynamic additions to the personality lexicon. *J Pers Soc Psychol* 2004;86:880-93.

51. Armeli S, Tennen H, Affleck G, et al. Does affect mediate the association between daily events and alcohol use? *J Stud Alcohol* 2000;61:862-71.

52. Stein KF. Affect instability in adults with a borderline personality disorder. *Arch Psychiatr Nurs* 1996;10:32-40.

53. Whalen CK, Jamner LD, Henker B, et al. The ADHD spectrum and everyday life: experience sampling of adolescent moods, activities, smoking, and drinking. *Child Dev* 2002;73:209-27.

54. Steiger H, Gauvin L, Jabalpurwala S, et al. Hypersensitivity to social interactions in bulimic syndromes: relationship to binge eating. *J Consult Clin Psychol* 1999;67:765-75.

55. Smyth J, Wonderlich S, Crosby R, et al. The use of ecological momentary assessment approaches in eating disorder research. *Int J Eat Disord* 2002;30:83-95.

56. Hillbrand M, Waite BM, Miller DS, et al. Serum cholesterol concentrations and mood states in violent psychiatric patients: an experience sampling study. *J Behav Med* 2000;23:519-29.

57. Peeters F, Nicholson NA, Berkhof J. Cortisol responses to daily events in major depressive disorder. *Psychosom Med* 2003;65:836-41.

58. Axelson DA, Bertocci MA, Lewin DS, et al. Measuring mood and complex behavior in natural environments: use of ecological momentary assessment in pediatric affective disorders. *J Child Adolesc Psychopharmacol* 2003;13:253-66.

59. Barge-Schaapveld DQ, Nicolson NA. Effects of antidepressant treatment on the quality of daily life: an experience sampling study. *J Clin Psychiatry* 2002;63:477-85.

60. 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.

61. Moskowitz DS, Pinard G, Zuroff DC, et al. The effect of tryptophan on social interaction in every day life: A placebo-controlled study. *Neuropsychopharmacology* 2001;25:277-89.

62. Moskowitz DS, Côté S. Do interpersonal traits predict affect? A comparison of three models. *J Pers Soc Psychol* 1995;69:915-24.

63. Moskowitz DS, Zuroff DC. Assessing interpersonal perceptions with the interpersonal grid. *Psychol Assess* 2005;17:218-30.

64. Moskowitz DS, Suh EJ, Desaulniers J. Situational influences on gender differences in agency and communion. *J Pers Soc Psychol* 1994;66:753-61.

65. Tse WS, Bond AJ. Serotonergic intervention affects both social dominance and affiliative behavior. *Psychopharmacology (Berl)* 2002;161:324-30.

66. Dubini A, Bosc M, Polin V. Do noradrenaline and serotonin differentially affect social motivation and behaviour? *Eur Neuropsychopharmacol* 1997;7(Suppl 1):S49-55.