



Published in final edited form as:

*Qual Life Res.* 2016 March ; 25(3): 497–506. doi:10.1007/s11136-015-1054-z.

## Ambulatory and diary methods can facilitate the measurement of Patient Reported Outcomes

Stefan Schneider, PhD and Arthur A. Stone, PhD

University of Southern California

### Abstract

**Purpose**—Ambulatory and diary methods of self-reported symptoms and wellbeing have received increasing interest in recent years. These methods are a valuable addition to traditional strategies for the assessment of Patient Reported Outcomes (PROs) in that they capture patients' recent symptom experiences repeatedly in their natural environments. In this article, we review ways that incorporating diary methods into PRO measurement can facilitate research on quality of life.

**Methods**—Several diary methods are currently available and they include “real time” (Ecological Momentary Assessment) and “near real time” (end-of-day assessments, Day Reconstruction Method) formats. We identify the key benefits of these methods for PRO research.

**Results**—(1) In validity testing, diary assessments can serve as a standard for evaluating the ecological validity and for identifying recall biases of PRO instruments with longer-term recall formats. (2) In research and clinical settings, diaries have the ability to closely capture variations and dynamic changes in quality of life that are difficult or not possible to obtain from traditional PRO assessments. (3) In test construction, repeated diary assessments can expand understanding of the measurement characteristics (e.g., reliability, dimensionality) of PROs in that parameters for differences between people can be compared with those for variation within people.

**Conclusions**—Diary assessment strategies can enrich the repertoire of PRO assessment tools and enhance the measurement of patients' quality of life.

### Keywords

Patient Reported Outcomes; Ambulatory Measurement; Diaries; Ecological Momentary Assessment; Day Reconstruction Method; Recall

---

Diary methods for the assessment of self-reported behaviors and experiences have received increasing interest in recent years as evidenced by articles published in highly regarded journals [1–6]. The family of ambulatory and diary methods encompasses “real time” diaries

---

Corresponding author: Stefan Schneider, Ph.D., Dornsife Center for Self-Report Science and Center for Social & Economic Research, University of Southern California, 635 Downey Way, Los Angeles, CA 90089-3332, schneids@usc.edu.

**Conflict of interest:** S.S. declares that he has no conflict of interest. A.A.S. is a Senior Scientist with the Gallup Organization and a Senior Consultant with ERT, inc.

### Compliance with Ethical Standards

**Ethical approval:** This article does not contain any studies with human participants or animals performed by any of the authors.

as realized in Experience Sampling or Ecological Momentary Assessment (EMA) techniques, as well as “near real time” formats including end-of-day assessments and the Day Reconstruction Method (DRM). Compared to more “traditional” assessment formats, which typically involve global or retrospective reports, diary methods are characterized by short recall periods and (often densely) repeated sampling of behaviors and experiences in people’s natural environments. These methods aim to reduce memory biases, increase the ecological validity of self-reports [7], and provide high resolution data that can capture the dynamic ebb and flow of people’s experiences [8,9].

Despite the burgeoning interest in diary assessments in many areas relevant to quality of life research, including psychopathology [10–12], emotional wellbeing [5,13,14], life span development [15,16], and behavioral medicine [17–19], attention to the potential benefits of diary techniques in the area of Patient Reported Outcomes (PROs) measurement has been limited. It is therefore timely to review how incorporating diary methods into the repertoire of PRO assessments can facilitate research on quality of life. To set the stage, we start by discussing the role of recall processes in patient self-reports, and provide an overview of different diary techniques for PRO assessment. We then review the utility of diary assessments for evaluating the ecological validity and potential recall biases of “traditional” retrospective PROs. Next, the usefulness of diaries for measuring temporal dynamics of quality of life is discussed, and we review strategies for developing PROs based on temporal patterns of experience. Finally, we highlight how densely repeated assessments with diaries can be used to expand testing of the measurement characteristics (e.g., reliability, dimensionality) of PROs from cross-sectional differences between individuals to variations and changes within individuals.

## Diary methods in PRO measurement: Rationale and options

How people access information from memory when completing self-reports is central to the rationale for employing diary methods. Research in cognitive psychology distinguishes between *semantic* and *episodic* memory as two mental processing systems involved in self-report [20]. Semantic memory represents conceptual knowledge that is not bound to specific contexts or episodes, including one’s judgements and beliefs about oneself. Episodic memory, on the other hand, represents information of one’s actual subjective *experiences* at particular times and situational contexts [20]. This distinction can be mapped onto most common PROs (see also Stull et al. [21] for a discussion of optimal reporting periods of PROs).

By design, some types of PRO measures tap *semantic* memory directly, such as measures of knowledge about chronic illness self-care regimens [22] or measures of health values and health preferences [23]. Other PROs discourage the use of episodic memory by asking respondents how they feel “in general”. Examples are the “global health” item from the 36-item Short Form Health Survey or the Patient-Reported Outcomes Measurement Information System (PROMIS<sup>®</sup>) global items (e.g., “In general, would you say your quality of life is” ... excellent – poor) [24]. Since no reporting period is specified in these items, respondents presumably access semantic knowledge in their ratings [20]. Similarly, physical functioning is commonly assessed by asking people whether they think that they *are able* to

perform certain behaviors (e.g., “are you able to exercise for an hour”) regardless of their actual behaviors [25]; this reporting format also likely evokes patients’ semantic knowledge about their behavior and functioning. For these PROs derived from semantic memory, we see little or no advantage to diary methods, unless one is specifically interested in how the reporting context might influence such ratings [26].

Much more common are PRO measures that ask people to report their actual behaviors and experiences over a particular time frame (e.g., the past week or the past month). The 36-item Short Form Health Survey, for example, asks people to recall their physical and emotional symptoms during the last 4-weeks, and also offers an “acute” 1-week recall version [27]. Items developed by PROMIS use a 7-day recall format for the majority of quality of life domains [28]. For retrospective reports to be accurate requires that respondents are able to access the necessary information accurately from *episodic* memory. However, the longer the reporting period, the less people can generally remember their actual contextual experiences and accessible memories become few and fuzzy. People may use mental shortcuts or draw on semantic knowledge in their ratings, which accentuates responses that are consistent with norms, stereotypes, and general beliefs [26,29,30].

Diary methods have much to offer for these latter types of PROs by using short reporting periods to facilitate accurate recall from episodic memory. The potential benefits for PRO assessment have been recognized, for example, when PROs are used as end-points to support drug label claims. The PRO guidance issued by the U.S. Food and Drug Administration states that “items with short recall periods or items that ask patients to describe their current or recent state are usually preferable.” [31, p.14] and diary methods have been used in many symptom domains.

Characteristics of different diary assessment strategies are summarized in Table 1. Among ambulatory diary methods, EMA has often been regarded a “gold standard” for measuring patients’ symptom experiences [9]. EMA has people describe their symptoms, thoughts, feelings, and/or behaviors as they are happening in real time (or close to real time), such that ratings are generated by respondents immediately as they access information from working memory. Early procedures collected momentary data using paper-and-pencil surveys sometimes combined with audible pagers [32], but advances in mobile and computer technology make more sophisticated options readily available, with EMA being implemented on electronic diaries and smartphones [32,33]. Combined with advances in sensor technology, this also provides increasing options for linking momentary PRO data with ambulatory measures of physiology (e.g., heart rate), physical activity (e.g., actigraphy), and geographical location [33,34].

Despite the advantages of EMA, it carries a cost for researcher and patient in that it is relatively expensive and involves significant time and burden for participants, which can impact the representativeness of those who are willing to participate in EMA studies and the moments that are captured with EMA [7,21,35]. End-of-day diaries, in which patients report their symptoms on a daily basis, often represent a viable alternative where EMA is not feasible or not desirable (e.g., when respondents are not able to respond to momentary prompting) [36–38]. End-of-day ratings require some degree of retrospection (typically, a

24-hr recall period is used). However, the amount of memory bias in end-of-day recall appears to be limited, at least for some symptoms. A recent study demonstrates this point: Broderick et al. [35] found near perfect correspondence between 1 week of EMA ratings and the average of 7 end-of-day ratings for symptoms of pain and fatigue.

Some research questions require the detail afforded by EMA and cannot be answered with end-of-day diaries, such as changes in PROs in response to specific social contexts and behaviors during the day (e.g., pain contingent upon activity)[9,39]. On the other hand, certain types of PROs would be intrusive to measure with EMA (e.g., measures of sleep disturbance) or require inferences about the impact of symptoms on people's lives (e.g., pain interference); such PROs are more easily assessed with daily diaries [36,40]. Importantly, end-of-day diaries are considerably less burdensome than EMA and offer the prospect of monitoring individuals for extended periods of time (up to many months).

The DRM offers yet another alternative to real-time momentary assessment [6,13]. This method has participants create a diary consisting of a sequence of *yesterday's* episodes, as is done in time-use studies. Revisiting the episodes of the previous day is intended to facilitate retrieval of contextual information from episodic memory [6]. Respondents then describe the episodes in terms of activities, social contexts, and rate their feelings during each episode. Increasing evidence suggests that DRM ratings replicate intra-day patterns of affect obtained with experience sampling methods [6,13] and can be used to study relationships between affect and physiological processes like ambulatory heart rate [2,41] and salivary cortisol [42]. The DRM can be administered with a questionnaire, over the Internet, or via an interview, making it attractive for addressing population-level research questions with relatively large samples [6,43]. To date, the DRM has not found widespread attention in PRO research, possibly because it has originally been developed to investigate relationships between time-use and emotions in general population samples. However, in theory, the principles underlying the DRM can be applied to the assessment of various PROs in large patient samples, including physical symptoms (e.g., pain, fatigue, chronic obstructive pulmonary disease symptoms) or chronic illness self-management behaviors (e.g., adherence to regimens across different contexts of the day). The full potential of the method for quality of life research remains to be explored. A drawback of the DRM is that reconstructing all episodes of the day is time intensive for participants, which makes it less feasible for applications that require PRO assessments to be brief (e.g., in clinical settings), and for consecutive assessment across many days (although shorter versions adapted from the DRM could potentially increase flexibility of administration [44]).

## Use of diaries for evaluating ecological validity and recall biases of PROs

It goes without saying that having valid PRO measures is essential for effective clinical research and practice. An often overlooked and unappreciated type of validity is *ecological validity*, and it indicates the degree to which study methods and findings approximate the real-life situation under investigation. When developing and validating PRO measures with longer recall periods, data obtained from diaries can serve an important function as a standard of comparison to test the ecological validity of recall-based PROs [7,45,46].

To date, several compelling findings have resulted from research on ecological validity. Studies comparing recall ratings with aggregated diary ratings over the same reporting period have usually found that symptoms tend to be recalled at higher levels relative to momentary and end-of-day ratings, a phenomenon sometimes referred to as the “experience memory gap” [47]. This effect appears to be robust especially for physical symptoms and negative affective experiences, which are generally rated as more intense, more frequent, and longer lasting in recall [38,40,48–51]. The magnitude of this “gap” between retrospective and diary ratings has been found to be similar for relatively shorter (7-day) and longer (30-day) reporting periods [45,36]. Frequency of behaviors also tends to be overestimated in recall [52,53], although some studies on alcohol use found the opposite pattern of lower frequencies in recall [54,55].

Apart from comparisons of mean symptom levels, studies have examined the agreement between recall ratings and diary ratings in correlational analyses with inconsistent results. While some studies have found reasonably high correspondence between aggregated momentary or daily assessments and recall ratings of a week or more [36,38,56–58], others suggest moderate to low correspondence for recall periods beyond one day [45,46,53]. A difficulty associated with tests of agreement between diaries and longer-term recall is that the magnitude of correlations required to judge the scores as sufficiently equivalent is not clear and may differ depending upon the research question [59]. For example, an intraclass correlation of .80 may be considered adequate for research purposes that involve comparisons between clinical groups, but may be insufficient in clinical settings where treatment decisions need to be based on precise data for individual patients.

A likely reason for the mixed results regarding the diary—recall correspondence is that the amount and direction of memory bias in recall ratings can be moderated by a variety of factors, including characteristics of the patients, measurement setting, and the nature of the symptom experiences. People tend to emphasize salient experiences and singular moments in recall, as suggested by the “peak-end” rule, and several studies have evidenced that recall ratings are disproportionately influenced by the highest (“peak”) and final (“end”) symptom levels of the reporting period [60–62]. Recall ratings can also be impacted by the emotional state at the time of symptom experience [63], by the amount of fluctuations or changes in experienced symptom levels [64–66], and by the symptom intensity at the time of recall [67]. Finally, recall accuracy can be moderated by people’s general beliefs and personality traits: McFarland et al. [29] demonstrated that women retrospectively overestimated their symptoms for menstrual days more strongly to the extent that this confirmed their beliefs about menstrual distress, and Feldman-Barrett [30] showed that retrospective emotion ratings were negatively biased for more neurotic individuals, whereas they were positively biased for more extraverted individuals.

An important implication of these findings is that conclusions about the degree of ecological validity may not translate from one study population, measurement context, and PRO measure to the next. As is the case for any other dimension of validity (e.g., factorial validity, convergent and discriminant validity), ecological validity is sample, situation, and instrument dependent. So far, we have only a fragmented picture of memory bias in recall PROs, and we suggest that diary methods can provide an essential resource for more

systematic ecological validity testing across different PRO domains, instruments, and patient populations.

## Use of diaries for capturing processes and temporal patterns of quality of life experiences

Ambulatory self-report measures also have the ability to capture unique types of information about patients' symptoms and quality of life that are difficult or impossible to obtain from traditional PROs. Diary measurement, and in particular momentary assessment, typically provides high resolution data from multiple assessments of the same construct, presenting the opportunity to study dynamic processes and how they unfold over time. Diaries have been used to closely monitor the course of symptoms following surgery [68,69], and to examine the exact course of physical and emotional side effects during chemotherapy [37,70,71]. They have also been employed in the investigation of situational contingencies of behaviors and experiences, such as the effect of momentary mood on smoking and substance use [72], emotional and physical symptom reactivity to daily stressors [73], and the effects of sleep disturbance on next-day symptoms [74].

Examination of diary data also makes apparent how much variability there is in patients' symptom experience. Traditional PRO measures often emphasize the stable "trait" aspect of people's quality of life; for example, an implicit assumption of self-report measures that ask about the "past 7 days" is that patients are best characterized by their average symptom level over that period. By contrast, diary studies show that there is substantial moment-to-moment and day-to-day variability in symptoms [75,76], and concepts that are commonly considered to be traits show surprisingly large short-term fluctuations within individuals [77,78]. While within-subject sources of variation are often regarded statistical nuisance and random measurement error, accumulating evidence suggests that symptom variability carries valuable information and is an important construct in itself [16,79].

A few examples demonstrate the importance of variability. Zautra et al. [80] examined the amount of variability in daily diary reports of fatigue among patients with rheumatoid arthritis, osteoarthritis, and fibromyalgia syndrome. Compared to arthritis patients, fibromyalgia patients showed significantly more pronounced daily fluctuations in fatigue, suggesting differences in the etiologies between these chronic pain conditions. Similarly, Trull et al. [81] showed greater momentary affect variability in borderline patients compared to patients with affective disorders, consistent with the notion that emotional instability represents a core diagnostic feature of borderline disorder. Differences in variability also have predictive value: longitudinal studies suggest that variability in cognitive performance is predictive of subsequent cognitive decline [82], and that greater variability in perceived life control predicts 5-year mortality [83].

Another application enabled by high resolution data is the extraction of rhythmic patterns from variability in people's diary ratings. A promising approach uses dynamical systems analysis (e.g., differential equation modeling) to understand cyclic changes in intensive momentary or daily time series [79]. These models often hypothesize that symptoms or emotions oscillate around a point of homeostasis, and individuals can be described by the



amplitude and frequency of oscillation and by the time it takes them to regulate their current state to their ideal set point [84,85]. Another cyclic pattern of within-person variability is produced by *diurnal patterns*. It is acknowledged that almost all biological processes have some diurnal rhythm [86], and a large-sample DRM study found different diurnal rhythms for specific affective and physical experiences [13]. Differences in diurnal pain patterns among patient subgroups have been linked to depression and somatization [87], which may have important implications for understanding biopsychosocial mechanisms underlying patients' quality of life.

We believe there is rich potential in ambulatory diary data for the construction of outcome measures. All too often, momentary or daily reports are simply averaged to create PRO measures representing a period of time, yet there are many ways in which diary reports can be aggregated to construct alternative PROs, some of which are schematically illustrated in Figure 1. Different indices derived from diaries could involve the frequency of acute shifts in symptom levels, the consistency of symptoms, the proportion of symptom ratings above a certain threshold, or the percent of time without symptoms (see [88,89], for review and conceptualization of various PRO measures that can be derived from EMA or daily diaries). Importantly, these new measures may capture aspects of symptom experience that are more meaningful to respondents, which in turn could enhance the use of PROs in clinical trials. It is possible that many new insights may emerge from alternative outcomes created from PRO diaries that deepen our understanding of adaptive or maladaptive symptom patterns, differences between illnesses, and clinically meaningful change in symptoms.

## Using diaries to evaluate PRO measurement characteristics within individuals

Contemporary PRO measurement construction has benefitted tremendously from recent developments in quantitative methodology, such as latent variable models based on confirmatory factor analysis and item response theory [90]. However, the psychometric evaluation of PRO measures still predominantly relies on cross-sectional, between-subjects designs, where data are gathered on many individuals, each of whom is measured on only one occasion. An often untested assumption is that the psychometric characteristics of a measure generalize from the level of between-person differences to the level of within-person change [91,92]. As pointed out by Molenaar [92] "only under very strict conditions ... can a generalization be made from a structure of interindividual variation to the analogous structure of intraindividual variation" (p. 201). Inappropriately inferring equality of measurement properties established for variation between people to longitudinal variation within an individual represents a form of *ecological fallacy* [93]. Recent psychometric advances using multilevel analysis of repeated diary assessments can help provide useful and novel insights in the psychometric characteristics of PROs in that they afford the comparison of measurement parameters obtained for between-person differences and those for within-person changes.

An appropriate PRO measure should have high reliability and minimal measurement error, but reliabilities established for cross-sectional data do not necessarily translate to those for within-person change processes. Cranford et al. [75] demonstrate an approach to assessing

the reliability of within-person changes based on multi-item daily diary measures. The approach is based on generalizability theory, whereby the variance of a set of scores is partitioned into multiple effects and their interactions (person by day by item). This allows for flexible reliability estimation of a measure used either in a between-person or a within-person design. Alternatively, multilevel confirmatory factor analysis procedures can be used to directly compare common reliability coefficients (Cronbach's alpha, Donald's omega) across inter- and intra-individual levels of analysis [94]. Diary studies have found reliabilities to be lower for within-person than for between-person levels of analysis [95,96], which may need to be considered in designs that require measures with high sensitivity to intra-individual change.

Factor analysis is commonly used to establish the underlying dimensionality of constructs measured by PRO items. Again, between- and within-person results on dimensionality can be different [91,97]. Roesch et al. [97] examined the within- and between-person factor structure of coping with everyday life stress from daily diary data using multilevel factor analysis. While some coping factors (Problem-Focused Coping and Social Support) emerged equally on both within-person and between-person levels, other factors (Minimization, Rumination, Avoidance, Distraction) were specific to the within-person or between-person levels but not both.

Even if the factors are the same for between- and within-person levels of analysis, it may still be the case that certain items are not equally sensitive to between-person differences and changes across time. Using repeated diary PRO assessments, this possibility can be examined by comparing the discrimination parameters (or factor loadings) of each item across between- and within-person levels in multilevel item response theory (or confirmatory factor analysis) models [91,98]. We previously applied this strategy to daily diary versions of PROMIS depression items and found that items tapping into depressive mood (feeling "sad", "depressed", "unhappy") discriminated day-to-day changes in depressive states significantly more strongly than they discriminated depression differences between people; this was not the case for items on cognitive aspects of depression (feeling "like a failure", "nothing to look forward to") [98]. These applications for diary PRO data may be useful for identifying constructs and items that are predominantly suitable for measuring stable or trait-like aspects of quality of life and discriminating them from items that are particularly sensitive to change and short-term fluctuations in symptoms and functioning.

Finally, a fundamental property of PRO measures is that respondents should interpret questions in a consistent manner, such that the measurement model is equivalently applicable to all people in the target population. Violations of this assumption are known as lack of measurement invariance or differential item functioning (DIF). Various techniques have been developed to examine DIF in cross-sectional data [99–101]. A limitation of most of these techniques is that the group characteristics for which DIF is suspected to be relevant must be known; this typically restricts DIF testing to a limited number of variables (e.g., demographic characteristics). Using real-time diary assessments, novel perspectives for DIF testing become available that go beyond the reach of traditional cross-sectional PRO assessments. For example, developments in multilevel latent variable modeling applied to



diary data have made possible to estimate individual-specific discrimination parameters in item response theory models by estimating factor loadings that are allowed to vary randomly across individuals [102,103]. In these models, variation in factor loadings captures the overall *degree* to which an item functions differentially across people, whereby the relevant group characteristics responsible for DIF do not need to be known. This could prove very fruitful for identifying items that are particularly prone to DIF.

As these examples illustrate, data collected repeatedly with EMA or daily diaries can address perspectives on the measurement properties of PROs that are often not possible within standard settings of PRO assessment. It should be emphasized that some of the associated statistical techniques, such as multilevel modeling extensions of latent variable models, are complex and their implementation requires careful consideration of many details. However, these new procedures are becoming more readily available for applied research as they are incorporated in commercial statistical analysis packages, and they enable important avenues for future PRO research.

## Conclusions

Diary techniques have a considerably long history in research and clinical practice, starting with written symptom records more than 50 years ago (see [104]). Technological and methodological advances, such as EMA implemented on smartphones, end-of-day diaries collected over the Internet, or DRM assessments administered to large-scale patient samples, now allow widespread adoption of these methods. Diary methods open windows into patients' symptom and wellbeing experiences in their natural daily environment, capturing patients' quality of life "as it is lived". They can reduce unsystematic and systematic errors from memory, provide unique insight into the processes and dynamics of symptom experience, offer new options for the development of outcome measures that capture respondents' experiences in meaningful ways, and enhance perspectives for a more comprehensive testing of measurement properties between and within individuals. Diary and traditional PRO assessment formats should not be viewed as competing options, but rather as complementing each other in PRO research and practice. Used in concert with other assessment formats, ambulatory and diary methods have the potential to enrich the repertoire of PRO assessment tools and to facilitate a comprehensive perspective on the measurement of quality of life.

## Acknowledgments

**Funding:** This work was supported by grants from the National Institute on Aging and the National Bureau of Economic Research, R01 AG042407-01A1, 5R01AG040629, and P01 AG05842.

## References

1. Stone AA, Shiffman S. Ecological Momentary Assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*. 1994; 16:199–202.
2. Bhattacharyya MR, Whitehead DL, Rakhit R, Steptoe A. Depressed mood, positive affect, and heart rate variability in patients with suspected coronary artery disease. *Psychosomatic Medicine*. 2008; 70:1020–1027. [PubMed: 18941130]

3. Stone AA, Schwartz JE, Broderick JE, Deaton A. A snapshot of the age distribution of psychological well-being in the United States. *Proceedings of the National Academy of Sciences*. 2010; 107:9985–9990.
4. Kahneman D, Deaton A. High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*. 2010; 107:16489–16493.
5. Steptoe A, Wardle J. Positive affect measured using ecological momentary assessment and survival in older men and women. *Proceedings of the National Academy of Sciences*. 2011; 108:18244–18248.
6. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: The day reconstruction method. *Science*. 2004; 306:1776–1780. [PubMed: 15576620]
7. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annual Review of Clinical Psychology*. 2008; 4:1–32.
8. Bolger N, Davis A, Rafaeli E. Diary methods: Capturing life as it is lived. *Annual review of psychology*. 2003; 54:579–616.
9. Stone, AA.; Shiffman, S.; Atienza, A.; Nebling, L., editors. *The science of real-time data capture: Self-reports in health research*. New York: Oxford University Press; 2007.
10. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological Assessment*. 2009; 21:463–475. [PubMed: 19947781]
11. aan het Rot M, Hogenelst K, Schoevers RA. Mood disorders in everyday life: A systematic review of experience sampling and ecological momentary assessment studies. *Clinical Psychology Review*. 2012; 32:510–523. [PubMed: 22721999]
12. Walz LC, Nauta MH, aan het Rot M. Experience sampling and ecological momentary assessment for studying the daily lives of patients with anxiety disorders: A systematic review. *Journal of anxiety disorders*. 2014; 28:925–937. [PubMed: 25445083]
13. Stone AA, Schwartz JE, Schkade D, Schwarz N, Krueger A, Kahneman D. A population approach to the study of emotion: diurnal rhythms of a working day examined with the Day Reconstruction Method. *Emotion*. 2006; 6:139–149. [PubMed: 16637757]
14. Reis HT, Sheldon KM, Gable SL, Roscoe J, Ryan RM. Daily well-being: The role of autonomy, competence, and relatedness. *Personality and Social Psychology Bulletin*. 2000; 26:419–435.
15. Carstensen LL, Pasupathi M, Mayr U, Nesselroade JR. Emotional experience in everyday life across the adult life span. *Journal of Personality and Social Psychology*. 2000; 79:644–655. [PubMed: 11045744]
16. Ram N, Gerstorf D. Time-Structured and Net Intraindividual Variability: Tools for Examining the Development of Dynamic Characteristics and Processes. *Psychology and Aging*. 2009; 24:778–791. [PubMed: 20025395]
17. Burton C, Weller D, Sharpe M. Are electronic diaries useful for symptoms research? A systematic review. *Journal of Psychosomatic Research*. 2007; 62:553–561. [PubMed: 17467410]
18. Conner TS, Barrett LF. Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine*. 2012; 74:327–337. [PubMed: 22582330]
19. Kubiak T, Stone AA. Ambulatory monitoring of biobehavioral processes in health and disease. *Psychosomatic Medicine*. 2012; 74:325–326. [PubMed: 22582329]
20. Robinson MD, Clore GL. Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychological bulletin*. 2002; 128:934–960. [PubMed: 12405138]
21. Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. *Current medical research and opinion*. 2009; 25:929–942. [PubMed: 19257798]
22. Dunn SM, Bryson JM, Hoskins PL, Alford JB, Handelsman DJ, Turtle JR. Development of the diabetes knowledge (DKN) scales: forms DKNA, DKNB, and DKNC. *Diabetes care*. 1984; 7:36–41. [PubMed: 6705664]
23. Craig BM, Busschbach JJ, Salomon JA. Modeling ranking, time trade-off and visual analogue scale values for EQ-5D health states: A review and comparison of methods. *Medical Care*. 2009; 47:634–641. [PubMed: 19433996]

24. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Quality of Life Research*. 2009; 18:873–880. [PubMed: 19543809]
25. Bruce B, Fries J. The health assessment questionnaire (HAQ). *Clinical and Experimental Rheumatology*. 2005; 23:S14–S18. [PubMed: 16273780]
26. Schwarz, N. Retrospective and concurrent self-reports: The rationale for real-time data capture. In: Stone, AA.; Shiffman, SS.; Atienza, A.; Nebeling, L., editors. *The science of real-time data capture: Self-reports in health research*. New York: Oxford University Press; 2007. p. 11–26.
27. Ware, JE. *SF-36 health survey: Manual & interpretive guide*. Boston: The Health Institute, New England Medical Center; 1993.
28. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of clinical epidemiology*. 2010; 63:1179–1194. [PubMed: 20685078]
29. McFarland C, Ross M, DeCourville N. Women’s theories of menstruation and biases in recall of menstrual symptoms. *Journal of Personality and Social Psychology*. 1989; 57:522–531. [PubMed: 2778636]
30. Barrett LF. The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personality and Social Psychology Bulletin*. 1997; 23:1100–1110.
31. U.S. Department of Health and Human Services Food and Drug Administration. [Accessed March 22, 2010] Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims. 2009. Available at <http://www.fda.gov>
32. Broderick JE. Electronic diaries: Appraisal and current status. *Pharmaceutical Medicine*. 2008; 22:69–74. [PubMed: 20037672]
33. Trull TJ, Ebner-Priemer U. Ambulatory assessment. *Annual Review of Clinical Psychology*. 2013; 9:151–176.
34. Bussmann, JB.; Ebner-Priemer, UW. Ambulatory assessment of movement behavior. In: Mehl, MR.; Conner, TS., editors. *Handbook of research methods for studying daily life*. New York, NY: The Guilford Press; 2012. p. 235–250.
35. Broderick JE, Schwartz JE, Schneider S, Stone AA. Can End-of-Day Reports Replace Momentary Assessment of Pain and Fatigue? *Journal of Pain*. 2009; 10:274–281. [PubMed: 19070550]
36. Broderick JE, Schneider S, Schwartz JE, Stone AA. Interference with activities due to pain and fatigue: accuracy of ratings across different reporting periods. *Quality of Life Research*. 2010; 19:1163–1170. [PubMed: 20535565]
37. Schlatter MC, Cameron LD. Emotional Suppression Tendencies as Predictors of Symptoms, Mood, and Coping Appraisals During AC Chemotherapy for Breast Cancer Treatment. *Annals of Behavioral Medicine*. 2010; 40:15–29. [PubMed: 20652465]
38. Junghaenel DU, Schneider S, Stone AA, Christodoulou C, Broderick JE. Ecological validity and clinical utility of Patient-Reported Outcomes Measurement Information System (PROMIS®) instruments for detecting premenstrual symptoms of depression, anger, and fatigue. *Journal of Psychosomatic Research*. 2014; 76:300–306. [PubMed: 24630180]
39. Murphy SL, Smith DM, Clauw DJ, Alexander NB. The impact of momentary pain and fatigue on physical activity in women with osteoarthritis. *Arthritis Care Res (Hoboken)*. 2008; 59:849–856.
40. Broderick JE, Schneider S, Junghaenel DU, Schwartz JE, Stone AA. Validity and reliability of patient-reported outcomes measurement information system (PROMIS) instruments in osteoarthritis. *Arthritis Care and Research*. 2013; 65:1625–1633. [PubMed: 23592494]
41. Daly M, Delaney L, Doran PP, Harmon C, MacLachlan M. Naturalistic monitoring of the affect-heart rate relationship: a day reconstruction study. *Health Psychology*. 2010; 29:186–195. [PubMed: 20230092]
42. Daly M, Baumeister RF, Delaney L, MacLachlan M. Self-control and its relation to emotions and psychobiology: evidence from a Day Reconstruction Method study. *Journal of Behavioral Medicine*. 2014; 37:81–93. [PubMed: 23160950]

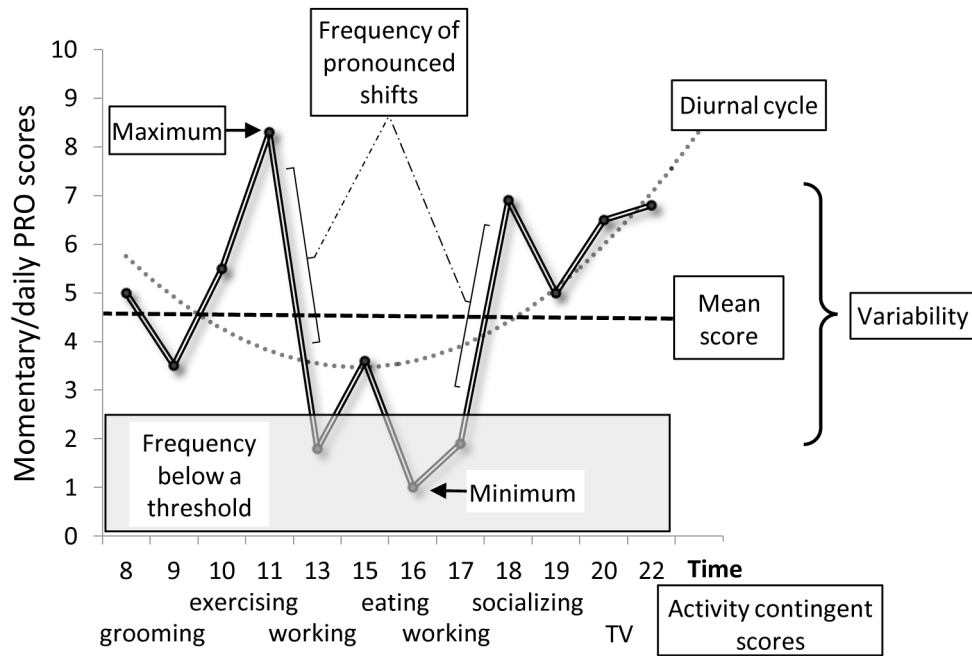
43. Dolan P, Kudrna L. More Years, Less Yawns: Fresh Evidence on Tiredness by Age and Other Factors. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2013;gbl118.
44. Smith J, Ryan LH, Queen TL, Becker S, Gonzalez R. Snapshots of Mixtures of Affective Experiences in a Day: Findings from the Health and Retirement Study. *Journal of population ageing*. 2014; 7:55–79. [PubMed: 24729799]
45. Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. *Pain*. 2008; 139:146–157. [PubMed: 18455312]
46. Stone AA, Broderick JE, Shiffman SS, Schwartz JE. Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain*. 2004; 107:61–69. [PubMed: 14715390]
47. Miron-Shatz T, Stone A, Kahneman D. Memories of yesterday's emotions: Does the valence of experience affect the memory-experience gap? *Emotion*. 2009; 9:885–891. [PubMed: 20001131]
48. Shrier LA, Shih MC, Beardslee WR. Affect and sexual behavior in adolescents: a review of the literature and comparison of momentary sampling with diary and retrospective self-report methods of measurement. *Pediatrics*. 2005; 115:e573–581. [PubMed: 15867022]
49. Bennett AV, Amtmann D, Diehr P, Patrick DL. Comparison of 7-day recall and daily diary reports of COPD symptoms and impacts. *Value in Health*. 2012; 15:466–474. [PubMed: 22583457]
50. Houtveen JH, Oei NY. Recall bias in reporting medically unexplained symptoms comes from semantic memory. *Journal of Psychosomatic Research*. 2007; 62:277–282. [PubMed: 17324676]
51. Van den Brink M, Bandell-Hoekstra E, Abu-Saad HH. The occurrence of recall bias in pediatric headache: a comparison of questionnaire and diary data. *Headache: The Journal of Head and Face Pain*. 2001; 41:11–20.
52. Homma Y, Ando T, Yoshida M, Kageyama S, Takei M, Kimoto K, et al. Voiding and incontinence frequencies: variability of diary data and required diary length. *Neurourology and urodynamics*. 2002; 21:204–209. [PubMed: 11948713]
53. Shiffman S. How many cigarettes did you smoke? Assessing cigarette consumption by global report, Time-Line Follow-Back, and ecological momentary assessment. *Health Psychology*. 2009; 28:519–526. [PubMed: 19751076]
54. Carney MA, Tennen H, Affleck G, del Boca FK, Kranzler HR. Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real-time. *Journal of Studies on Alcohol and Drugs*. 1998; 59:447–454.
55. Litt MD, Cooney NL, Morse P. Ecological momentary assessment (EMA) with treated alcoholics: methodological problems and potential solutions. *Health Psychology*. 1998; 17:48–52. [PubMed: 9459069]
56. Bennett A, Patrick D, Bushnell D, Chiou C, Diehr P. Comparison of 7-day and repeated 24-h recall of type 2 diabetes. *Quality of Life Research*. 2011; 20:769–777. [PubMed: 21088912]
57. Jamison RN, Raymond SA, Slawsby EA, McHugo GJ, Baird JC. Pain assessment in patients with low back pain: comparison of weekly recall and momentary electronic data. *Journal of Pain*. 2006; 7:192–199. [PubMed: 16516825]
58. Stel VS, Smit JH, Pluijm SMF, Visser M, Deeg DJH, Lips P. Comparison of the LASA Physical Activity Questionnaire with a 7-day diary and pedometer. *Journal of clinical epidemiology*. 2004; 57:252–258. [PubMed: 15066685]
59. Cella D, Stone AA. Health-related quality of life measurement in oncology: Advances and opportunities. *American Psychologist*. 2015; 70:175–185. [PubMed: 25730723]
60. Bogaerts K, Wan L, Van Diest I, Stans L, Decramer M, Van den Bergh O. Peak-end memory bias in laboratory-induced dyspnea: a comparison of patients with medically unexplained symptoms and healthy controls. *Psychosomatic Medicine*. 2012; 74:974–981. [PubMed: 23115343]
61. Stone AA, Broderick JE, Kaell AT, DelesPaul PAEG, Porter LE. Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritis? *Journal of Pain*. 2000; 1:212–217. [PubMed: 14622620]
62. Redelmeier DA, Katz J, Kahneman D. Memories of colonoscopy: a randomized trial. *Pain*. 2003; 104:187–194. [PubMed: 12855328]

63. Kent G. Memory of dental pain. *Pain*. 1985; 21:187–194. [PubMed: 3982842]
64. Stone AA, Schwartz JE, Broderick JE, Shiffman SS. Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. *Pers Soc Psychol Bull*. 2005; 31:1340–1346. [PubMed: 16143666]
65. Sohl SJ, Friedberg F. Memory for fatigue in chronic fatigue syndrome: relationships to fatigue variability, catastrophizing, and negative affect. *Behavioral Medicine*. 2008; 34:29–38. [PubMed: 18400687]
66. Schneider S, Broderick JE, Junghaenel DU, Schwartz JE, Stone AA. Temporal trends in symptom experience predict the accuracy of recall PROs. *Journal of Psychosomatic Research*. 2013; 75:160–166. [PubMed: 23915773]
67. Smith WB, Safer MA. Effects of present pain level on recall of chronic pain and medication use. *Pain*. 1993; 55:355–361. [PubMed: 8121697]
68. Begg A, Drummond G, Tiplady B. Assessment of postsurgical recovery after discharge using a pen computer diary. *Anaesthesia*. 2003; 58:1101–1105. [PubMed: 14616597]
69. Chapman CR, Davis J, Donaldson GW, Naylor J, Winchester D. Postoperative Pain Trajectories in Chronic Pain Patients Undergoing Surgery: The Effects of Chronic Opioid Pharmacotherapy on Acute Pain. *Journal of Pain*. 2011; 12:1240–1246. [PubMed: 22036517]
70. Jim HS, Small B, Faul LA, Franzen J, Apte S, Jacobsen PB. Fatigue, depression, sleep, and activity during chemotherapy: daily and intraday variation and relationships among symptom changes. *Annals of Behavioral Medicine*. 2011; 42:321–333. [PubMed: 21785899]
71. Junghaenel DU, Cohen J, Schneider S, Neerukonda AR, Broderick JE. Identification of distinct fatigue trajectories in patients with breast cancer undergoing adjuvant chemotherapy. *Supportive Care in Cancer*. 2015:1–9.
72. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. *Psychological Assessment*. 2009; 21:486–497. [PubMed: 19947783]
73. Almeida DM. Resilience and vulnerability to daily stressors assessed via diary methods. *Current Directions in Psychological Science*. 2005; 14:62–68.
74. Edwards RR, Almeida DM, Klick B, Haythornthwaite JA, Smith MT. Duration of sleep contributes to next-day pain report in the general population. *Pain*. 2008; 137:202–207. [PubMed: 18434020]
75. Cranford JA, Shrout PE, Iida M, Rafaeli E, Yip T, Bolger N. A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin*. 2006; 32:917–929. [PubMed: 16738025]
76. Schneider S, Junghaenel DU, Keefe FJ, Schwartz JE, Stone AA, Broderick JE. Individual differences in the day-to-day variability of pain, fatigue, and well-being in patients with rheumatic disease: Associations with psychological variables. *Pain*. 2012; 153:813–822. [PubMed: 22349917]
77. Barge-Schaapveld DQ, Nicolson NA, Berkhof J. Quality of life in depression: daily life determinants and variability. *Psychiatry Research*. 1999; 88:173–189. [PubMed: 10622339]
78. Maes IH, Delespaul PA, Peters ML, White MP, van Horn Y, Schruers K, et al. Measuring Health-Related Quality of Life by Experiences: The Experience Sampling Method. *Value in Health*. 2015; 18:44–51. [PubMed: 25595233]
79. Nesselroade JR, Ram N. Studying intraindividual variability: what we have learned that will help us understand lives in context. *Research in Human Development*. 2004; 1:9–29.
80. Zautra AJ, Fasman R, Parish BP, Davis MC. Daily fatigue in women with osteoarthritis, rheumatoid arthritis, and fibromyalgia. *Pain*. 2007; 128:128–135. [PubMed: 17055648]
81. Trull TJ, Solhan MB, Tragesser SL, Jahng S, Wood PK, Piasecki TM, et al. Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology*. 2008; 117:647–661. [PubMed: 18729616]
82. MacDonald SW, Hultsch DF, Dixon RA. Performance variability is related to change in cognition: evidence from the Victoria Longitudinal Study. *Psychology and Aging*. 2003; 18:510–523. [PubMed: 14518812]



83. Eizenman DR, Nesselroade JR, Featherman DL, Rowe JW. Intraindividual variability in perceived control in a older sample: The MacArthur successful aging studies. *Psychology and Aging*. 1997; 12:489–502. [PubMed: 9308096]
84. Chow SM, Ram N, Boker SM, Fujita F, Clore G. Emotion as a thermostat: representing emotion regulation using a damped oscillator model. *Emotion*. 2005; 5:208–225. [PubMed: 15982086]
85. Boker SM, Nesselroade JR. A method for modeling the intrinsic dynamics of intraindividual variability: Recovering the parameters of simulated oscillators in multi-wave panel data. *Multivariate Behavioral Research*. 2002; 37:127–160. [PubMed: 26824172]
86. Smolensky MH, D'Alonzo GE. Medical chronobiology: concepts and applications. *American Review of Respiratory Disease*. 1993; 147:S2–S19. [PubMed: 8494195]
87. Jamison RN, Brown GK. Validation of hourly pain intensity profiles with chronic pain patients. *Pain*. 1991; 45:123–128. [PubMed: 1876417]
88. Stone AA, Broderick JE, Schneider S, Schwartz JE. Expanding options for developing outcome measures from momentary assessment data. *Psychosomatic Medicine*. 2012; 74:387–397. [PubMed: 22582336]
89. Affleck G, Tennen H, Urrows S, Higgins P. Individual differences in the day-to-day experience of chronic pain: a prospective daily study of rheumatoid arthritis patients. *Health Psychol*. 1991; 10:419–426. [PubMed: 1765037]
90. Fayers, P.; Machin, D. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. Wichester, West Sussex: John Wiley & Sons; 2013.
91. Mehta PD, Neale MC. People are variables too: multilevel structural equations modeling. *Psychological Methods*. 2005; 10:259–284. [PubMed: 16221028]
92. Molenaar PC. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*. 2004; 2:201–218.
93. Velicer, WF.; Babbitt, SF.; Palumbo, R. Idiographic applications: Issues of ergodicity and generalizability. In: Molenaar, PCM.; Lerner, RM.; Newell, KM., editors. *Handbook of Developmental Systems Theory and Methodology*. New York, NY: Guilford Press; 2014. p. 425-441.
94. Geldhof GJ, Preacher KJ, Zyphur MJ. Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*. 2014; 19:72–91. [PubMed: 23646988]
95. Mäkikangas A, Kinnunen S, Rantanen J, Mauno S, Tolvanen A, Bakker AB. Association between vigor and exhaustion during the workweek: a person-centered approach to daily assessments. *Anxiety, Stress, & Coping*. 2014; 27:555–575.
96. Church AT, Katigbak MS, Ching CM, Zhang H, Shen J, Arias RM, et al. Within-individual variability in self-concepts and personality states: Applying density distribution and situation-behavior approaches across cultures. *J Res Pers*. 2013; 47:922–935.
97. Roesch SC, Aldridge AA, Stocking SN, Villodas F, Leung Q, Bartley CE, et al. Multilevel factor analysis and structural equation modeling of daily diary coping data: Modeling trait and state variation. *Multivariate Behavioral Research*. 2010; 45:767–789. [PubMed: 21399732]
98. Schneider S, Choi SW, Junghaenel DU, Schwartz JE, Stone AA. Psychometric characteristics of daily diaries for the Patient-Reported Outcomes Measurement Information System (PROMIS®): a preliminary investigation. *Quality of Life Research*. 2013; 22:1859–1869. [PubMed: 23180166]
99. Crane PK, Gibbons LE, Jolley L, van Belle G. Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Medical Care*. 2006; 44:S115–S123. [PubMed: 17060818]
100. Roussos L, Stout W. A multidimensionality-based DIF analysis paradigm. *Applied Psychological Measurement*. 1996; 20:355–371.
101. Zumbo BD. Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language assessment quarterly*. 2007; 4:223–233.
102. Asparouhov, T.; Muthén, B. General random effect latent variable modeling: Random subjects, items, contexts, and parameters. Paper presented at the annual meeting of the National Council on Measurement in Education; 2012.
103. De Boeck P. Random item IRT models. *Psychometrika*. 2008; 73:533–559.
104. Verbrugge LM. Health diaries. *Medical Care*. 1980; 18:73–95. [PubMed: 6986517]





**Figure 1.** Schematic illustration exemplifying different ways in which repeated diary reports can be summarized to construct alternative patient-reported outcome measures.

**Table 1**

Comparison of different types of ambulatory and diary PRO assessment strategies

Strategy	Reporting period	Characteristics
Ecological Momentary Assessment (EMA)	Moments or very short time periods	<ul style="list-style-type: none"> <li>• Least amount of recall bias, collected in patients' natural environment</li> <li>• Most fine-grained data, can capture within-day changes and detailed contextual information</li> <li>• Can be coupled with other real-time ambulatory measures (e.g., heart rate, actigraphy)</li> </ul>
End-of-day reports	Typically 24 hours	<ul style="list-style-type: none"> <li>• Less participant burden compared to EMA; allows monitoring people for longer time periods</li> <li>• Less intrusive than EMA (e.g., for PROs of sleep quality or sexual functioning)</li> <li>• Useful for PRO domains that involve evaluation of symptom impact (e.g., pain interference)</li> </ul>
Day Reconstruction Method (DRM)	Episodes of the day	<ul style="list-style-type: none"> <li>• Can be administered with large samples to address population-level questions (though significant time burden)</li> <li>• Allows examining linkages of episodic experiences with daily contexts and activities</li> <li>• Less intrusive than EMA, but typically one-time assessment of a single day</li> </ul>