

# Paper and Plastic in Daily Diary Research: Comment on Green, Rafaeli, Bolger, Shrout, and Reis (2006)

Howard Tennen and Glenn Affleck  
University of Connecticut School of Medicine

James C. Coyne  
University of Pennsylvania School of Medicine

Randy J. Larsen  
Washington University of St. Louis

Anita DeLongis  
University of British Columbia

The authors applaud A. S. Green, E. Rafaeli, N. Bolger, P. E. Shrout, and H. T. Reis's (2006) response to one-sided comparisons of paper versus electronic (plastic) diary methods and hope that it will stimulate more balanced considerations of the issues involved. The authors begin by highlighting areas of agreement and disagreement with Green et al. The authors review briefly the broader literature that has compared paper and plastic diaries, noting how recent comparisons have relied on study designs and methods that favor investigators' allegiances. The authors note some sorely needed data for the evaluation of the implications of paper versus plastic for the internal and external validity of research. To facilitate evaluation of the existing literature and assist in the design of future studies, the authors offer a balanced comparison of paper and electronic diary methods across a range of applications. Finally, the authors propose 2 study designs that offer fair comparisons of paper and plastic diary methods.

*Keywords:* diary methods, daily diary research, paper vs. electronic diaries

Controversy over “paper or plastic” has been played out in several venues. In the debate over whether consumers should request paper or plastic bags for their groceries, the Film and Bag Federation has accentuated the virtues of plastic. Likewise, in the discussion over using “paper or plastic” to pay for purchases, Discover Card has touted the advantages of credit cards, despite evidence that people often spend considerably more when they use plastic instead

of cash. A recent spate of reports advocating electronic diaries (EDs) over paper diaries (PDs) in psychological research—plastic over paper—has stirred controversy within the scientific community. Green, Rafaeli, Bolger, Shrout, and Reis's (2006) article is therefore a particularly welcomed methodological contribution. It could become an impetus for a more balanced discussion of the relative strengths and weaknesses of electronic and paper methods. In this commentary, we demonstrate how these strengths and weaknesses may have been exaggerated by advocates of one or the other diary method, and we attempt to offer a framework for a balanced appraisal of each approach.

---

Howard Tennen and Glenn Affleck, Department of Community Medicine and Health Care, University of Connecticut School of Medicine; James C. Coyne, Department of Psychiatry, University of Pennsylvania School of Medicine; Randy J. Larsen, Department of Psychology, Washington University of St. Louis; Anita DeLongis, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada.

National Institute of Mental Health Grant RO1-MH63732, National Institute of Alcohol Abuse and Alcoholism Grant P50-AA03510, Arthritis Foundation Grant 6-33998, and a grant from the Social Science and Humanities Research Council of Canada supported preparation of this article. We thank Ruth Tennen, Melady Preece, Dayna Lee-Baggeley, Susan Holtzman, and Tamlin Conner for their contributions to this commentary.

Correspondence concerning this article should be addressed to Howard Tennen, Department of Community Medicine and Health Care, University of Connecticut Health Center, Farmington, CT 06030-6325. E-mail: Tennen@nso1.uhc.edu

## Agreed: Participant Motivation Is Key in Diary Studies

Green et al.'s (2006) studies underscore the importance of participant motivation and investigators' efforts to achieve a collaborative engagement for the success of diary studies. The demands placed on diary study participants require that they feel involved in the research (Bolger, Davis, & Rafaeli, 2003), and Green et al. were careful to prepare participants for daily reporting and engage them fully in the research (see also DeLongis, Capreol, Holtzman, O'Brien, & Campbell, 2004; Rusting & Larsen, 1998). Working in close collaboration with research participants has also been en-

dorsed by advocates of plastic diaries (e.g., Stone & Shiffman, 2002). Fifteen years ago, Stone, Kessler, and Haythornthwaite (1991) recommended that diary researchers even warn potential participants how burdensome the task may be. They acknowledged the difficult compromises in internal and external validity that come with recruiting highly motivated individuals to participate in complex diary studies. Investigators who support the use of PDs and those who advocate EDs can find common ground in Green et al.'s demonstration of the importance of collaborative engagement with daily diary research participants.

### Where We Part Ways With Green et al. (2006)

Our enthusiasm for Green et al.'s (2006) comparison of PDs and EDs is dampened somewhat by what seems to be their use of a methodological mirror image of the studies that have underscored the advantages of EDs. A series of previous investigations (e.g., Stone, Shiffman, Schwartz, Broderick, & Hufford, 2003) compared typical PDs to best practice EDs, thus introducing a number of confounds that give a distinct advantage to EDs. Because Green et al.'s Studies 2 and 3 were originally conducted some time ago, they did not include many of the compliance- and response-enhancing features that today can be built into EDs. These EDs are more like plastic-encased questionnaires accompanied by a beeper than the current state-of-the-art ED.<sup>1</sup>

Green et al.'s (2006) efforts to cultivate their participants' cooperation involved having participants only try to enter responses on schedule, while encouraging them to enter even off-schedule responses. This may have engendered more noncompliance with both methods than would have been the case had there been an emphasis on strict compliance. The ability of EDs to ensure that responses can be entered only on schedule is an essential feature—the absence of which in Green et al.'s protocol weakened their comparisons of data equivalence. We are not suggesting that in daily diary studies, protocol compliance is more important than accurate responding. However, if fleeting experiences are the target of diary reports, high compliance rates are essential.

### New Plastic Versus the Same Old Paper

Whereas PDs will probably remain essentially unchanged, EDs are rapidly becoming more flexible and are thus able to help investigators address complex questions that remain beyond the capabilities of PDs. GPS receivers connected to cell phones will soon track individuals. Bluetooth technology will permit EDs to communicate with each other so that a participant's ED will prompt him or her to respond to specific diary queries when his or her partner, also a study participant, is nearby. ED technology has been applied to the detection of intoxication (e.g., Levit, Huber,

Batliner, & Noeth, 2001) and can now deliver short-term memory tests, record reaction time, administer measures of response inhibition, and administer a portable Implicit Association Test in situ (Dabbs, Bassett, & Dyomina, 2003).

EDs' ability to create and verify precisely equal intervals between diary entries is a critical advantage when the time series requires equal intervals between responses. As cell phones and PDA technologies begin to merge, keypad responses, text messaging, and voice response will be used in the same study. As people across socioeconomic categories are as likely to carry a cell phone as they are to carry a wallet, ED data entry portals may soon be viewed by participants as less intrusive than a pocket-sized PD. For all of these reasons, any comparison of PDs and EDs needs to consider their rapidly diverging capabilities.

### The Broader Literature: Paper and Plastic in Context

Whereas relatively few studies comparing PDs and EDs or questioning the validity of PDs have appeared in the psychological literature (e.g., Broderick, Schwartz, Shiffman, Hufford, & Stone, 2003; Gable, Reis, & Eliot, 2000; Litt, Cooney, & Morse, 1998), the relative merits of PD and ED reports have become a major focus in the medical and pharmacological literatures. J. M. Weiler (2005) recently reviewed over 20 such comparisons. Unlike Green et al.'s (2006) studies, which were for the most part secondary analyses of existing data, the studies reviewed by J. M. Weiler (2005) were designed explicitly to compare PDs and EDs across various medical conditions, and they paint a remarkably mixed picture of the relative merits of PDs and EDs. McKenzie et al. (2004) concluded that EDs provide data of superior scientific quality and integrity. In contrast, K. Weiler, Christ, Woodworth, Weiler, and Weiler (2004) deduced that data from PDs are indistinguishable from ED data, and Carpenter and Monahan (2004) inferred that both PDs and EDs may have seriously underestimated their target event (hot flushes in breast cancer survivors) compared with a skin-conductance monitoring device. Thus, some comparisons find that PDs and EDs yield data of comparable quality, others find an advantage to EDs, and still others find neither PDs nor EDs provide data of sufficient accuracy to serve as a measure of outcome in clinical trials.

Green et al.'s (2006) findings, as well as the assertions of investigators eager to espouse the superiority of EDs, must be considered in the context of this broader literature. One limitation of PDs that is acknowledged even by advocates of them is that to be effective, PDs, like EDs, require time-verified participant reports. Except in highly restricted en-

<sup>1</sup> As Green et al. (2006) explained in their article, their studies represent secondary analyses of data collected previously. Their EDs may have represented the state of the art in electronic data capture at the time the data were originally collected.

vironments, time-verified PDs are feasible only for once-a-day reports. The literature reviewed by J. M. Weiler (2005) suggests that the previous comparisons of PDs and EDs best known to behavioral investigators have not been fair comparisons and that Green et al.'s comparison shares this bias, albeit in the direction of supporting PDs.

### Unfair Comparisons Revisited

In their article "Unfair Comparisons," Cooper and Richardson (1986) asserted that for a comparison of theories, factors, or variables—and we would add data collection methods—to be fair, the competing methods must be implemented with equal care and fidelity, that is, they must achieve procedural equivalence. They suggested having independent judges determine whether the procedures provide an equal opportunity for supporting the competing methods.

Previous comparisons of PDs and EDs have not involved procedural equivalence and instead favored EDs, particularly when PDs have been applied to questions that would be better answered through EDs (on the basis of criteria we elaborate later in this commentary) and when verification of diary reports has been surreptitious. Studies comparing the best EDs with the worst PDs—that is, with no effort to time stamp the paper versions or to encourage participants to comply with the diary protocol (Stone & Shiffman, 2002)—are the clearest examples of unfair comparisons. Transparently monitoring compliance enhances compliance for both PDs and EDs. Yet in the most widely cited PD–ED comparisons (e.g., Broderick et al., 2003), respondents could conclude only that compliance would not be tracked when in fact compliance was monitored covertly. By comparing EDs with transparent compliance-monitoring capabilities to PDs designed to appear incapable of monitoring compliance, previous investigations have confounded inherent differences between PDs and EDs with participants' expectations about the demands of the study.

### Compliance Enhancement, Event-Contingent Recording (ECR), and Unfair Comparisons: The Devil in the Details

Two features of EDs that make them particularly attractive to investigators are their capacity to enhance compliance and their ability to record behavior or experiences in real time through ECR. Closer inspection of these ED features, however, demonstrates how they may contribute to unfair PD–ED comparisons.

### *ED Compliance-Enhancing Features Contribute to Unfair Comparisons*

EDs give respondents some degree of personal control over when they answer a scheduled diary entry by allowing

them to delay their response for a brief period and to defer an audible prompt when it would be inconvenient (e.g., in a church or theater). Although Stone and Shiffman (2002) have cautioned investigators about the risks associated with participant-initiated suspension of ED audible prompts and suggested that the frequency of such suspensions should be reported, we have not been able to locate extensive documentation regarding participants' use of these ED features. Yet, consider data from a study of women with fibromyalgia who were audibly prompted to provide ED reports four times a day for 30 days (Affleck et al., 2001); Affleck et al. (2001) reported that this data set included only 1.3% missing-data entries. More recently, however, it was noted that approximately 10% of the completed ED interviews in this study were participant-initiated delayed entries and that half of these had been delayed for 15 min after participants were prompted to reply (Tennen, Affleck, & Zautra, in press). In other words, ED compliance-enhancement features may lead to underestimates of noncompliance, depending on how long participants are allowed to delay responding and how strictly compliance is defined. When noncompliance is defined as diary reports that are not on time, a comparison between surreptitiously monitored PDs and EDs with compliance-enhancement features (e.g., Stone et al., 2003) will exaggerate any difference in compliance: An incorrectly timed response in a PD is evidence of noncompliance, whereas a delayed ED response is on time. This is an unfair comparison.

### *ED ECRs Contribute to Unfair Comparisons*

Investigators who champion EDs and challenge the validity of PDs rely heavily on ECRs to capture behavior in real time.<sup>2</sup> Yet, as Hufford, Shields, Shiffman, Paty, and Balabanis (2002) acknowledged, ECRs are not verifiable. They are just as vulnerable to backfilling as PDs, and missed ECRs cannot be detected. Thus, when Collins et al. (1998) stated that "87% of subject-initiated drinking assessments were started within 1 minute of starting to drink" (p. 310) they were referring to participants' reports that they had initiated these ECRs in a timely manner. Investigators who steadfastly endorse ECRs yet criticize PDs with postmarked verification are inexplicably holding a double standard.

<sup>2</sup> Whereas studies in the ecological momentary assessment literature (e.g., Hufford et al., 2002) appear to demonstrate that participant reactivity is modest, Silvia (2002) noted that tasks that cue increased self-awareness of one's own current affective state may reduce affect intensity. Other evidence suggests a direct association between number of measurement occasions per day and level of reactivity (Korotitsch & Nelson-Gray, 1999). The potential for reactivity with both signal- and event-contingent data collection warrants ongoing attention.

### When Paper Matches Plastic

Many important research questions can be addressed using once-daily PDs. Although ED advocates highlight the imprecision of PD verification methods, we believe that PDs can address the following questions as well as EDs.

#### *When Same-Day Relationships Are Informative*

For some research questions, same-day associations reliably derived from PDs can advance an area of inquiry in important ways (Cohen, Gunthert, Butler, O'Neill, & Tolpin, 2005; Tennen, Affleck, & Armeli, 2005). An investigator interested in the contextual determinants of risky sex among college students might hypothesize that days experienced as more stressful are more likely than less stressful days to be associated with risky sexual behavior. Of course, engaging in risky sex might generate concerns that make the day seem more stressful in retrospect, so causal inference may be limited. But as an initial pass at this question, a same-day association is informative, as is its absence.

#### *When the Behaviors Being Studied Are Discrete*

Both PD and ED advocates seem to agree that end-of-day reports using PDs minimize participant burden and are therefore useful in studies lasting several weeks or months. They also seem to agree that discrete or dichotomous occurrences, such as a headache during the day, can be captured reliably through once-daily PDs (e.g., Bolger, Davis, & Rafaeli, 2003; Stone & Shiffman, 2002). In the same study of determinants of risky sex described above, the investigator might predict that individuals are more likely to engage in risky sex on days when alcohol is consumed. Participants can presumably recall whether they consumed alcohol during the day or evening, whether they had sex, and whether they used a condom. These reports can be made even the next day, verified by postmark.<sup>3</sup> We enjoyed contemplating how an ED protocol might attempt to assess risky sex in real time and the potential bias in the kinds of research participants willing to provide such data. For some discrete behaviors, such as medication adherence, end-of-day reports mailed to investigators the next day are quite concordant with adherence measured via an electronic medication-event monitoring system (Feinn, Tennen, Cramer, & Kranzler, 2003).

#### *When Theory Predicts a Cross-Day Lagged Effect or Weekday–Weekend Cycles*

A cross-day lagged effect is also amenable to once-daily PDs. The same investigator depicted previously might predict that an evening that includes risky sex is more likely to be followed by heavier drinking the next day. This prediction is readily testable using postmarked once-daily PDs.

Another lagged effect is how weekday experience predicts the subsequent weekend's experience and behavior. We suspect that many individuals drink more on weekends that follow particularly stressful weeks. Any such sequencing can be examined reliably with PDs.

#### *When Daily Dependent Measures Include Future Behavior*

Many diary studies are designed to predict one day's behavior from the previous days' experience. A pain researcher, for example, may hypothesize that a day described at day's end as particularly painful may be followed by a day of more restrained activity. Wirtz, Kruger, Scallan, and Diener (2003) and Redelmeier, Katz, and Kahneman (2003) demonstrated that recall of one's experience can be a better predictor of future behavior than experience measured online through an ED, and they encouraged investigators not to ignore recalled experience. It would therefore be unfortunate if investigators retreated from measuring the recollection of a day's experience via PD and instead relied solely on ED experience sampling methods or ecological momentary assessment.

#### When Plastic Is the Method of Choice

There are research questions that cannot be addressed reliably though PDs and therefore require EDs. These include hypothesized within-day temporal dynamics or a focus on aspects of daily experience that may be fleeting and not well recalled later in the day.

#### *When Examining Within-Day Temporal Dynamics*

Many behavioral phenomena unfold within a single day. A once-daily diary, paper or plastic, simply cannot capture the temporal dynamics of such phenomena. Consider, for example, the self-medication hypothesis, which asserts that vulnerable individuals turn to alcohol to dampen painful emotions and that the relief alcohol brings reinforces continued self-medication. An adequate test of this hypothesis requires the temporal sequence of emotional experience and alcohol use within each day sampled over many days. For

<sup>3</sup> We acknowledge that postmarks are a less than ideal method for time stamping participants' diary reports. We consider postmarks both a "bogus pipeline" to enhance adherence and a way to document that participants are not hoarding diary reports. Indeed, in six separate studies, we found almost no evidence of hoarding. In a recent unpublished daily diary study (Barta, Kiene, Tennen, & Abu-Hasaballah, 2006) of individuals living in areas with high crime rates, where mail theft is not uncommon, we found that a telephone-based interactive voice-response method yielded a high rate of verified diary entries. Overall, postmarks on PDs are imperfect but for many purposes adequate.

vulnerable individuals (e.g., people with a family history of alcohol dependence), emotional distress during the day should be followed by greater alcohol use in the evening, and alcohol use in the evening should lead to reduced distress. Such hypotheses cannot be addressed adequately with PDs because of the requirement for verified within-day temporal sequencing of the variables under study (cf. Swendsen et al., 2000).

### *When Evaluating Rapidly Changing Phenomena*

Rapidly changing phenomena, such as emotions, can be assessed reliably only when measured close to their real-time occurrence. People seem unable to reliably recall their coping efforts after a few days (Ptacek, Smith, Espe, & Raffety, 1994; Stone et al., 1998), though it is not clear whether they can accurately recall coping strategies at day's end. Although end-of-day recollections of emotional experiences are vulnerable to the retrieval biases that affect episodic memory (Stone, Shiffman, & DeVries, 1999) and are likely to overestimate real-time affective reactions (Barrett, 1997), same-day recollections show good accuracy when compared with online reports (Thomas & Diener, 1990; Tugade, Conner, & Barrett, in press).

However, even the staunchest ED advocates have not provided guidelines regarding the appropriate frequency or duration of ED reports to address various questions, and most ED protocols seem to be based on issues related to participant burden and ED capabilities rather than informed speculation about the timing of assessments for the phenomenon being investigated. A critical aim of daily diary data is to represent a universe of occasions, just as cross-sectional data aim to represent a population of individuals. If the behavior is relatively rare, sampling at random moments during the day is inefficient and likely to miss the behavior of interest. Such behaviors can be recorded reliably at day's end or through ECR, both of which are equally reliable as PD or ED indicators. If the process of interest is presumed to be cyclical, the sampling rate would need to be at least twice the rate of the periodicity in the phenomenon to avoid a problem known as aliasing in the time-series literature. Only rarely have data collection schedules in PD or ED studies been based on how the phenomenon under study is thought to unfold temporally.

### *When Paper May Be More Effective Than Plastic*

Bolger et al. (2003) summarized several difficulties associated with EDs that may make PDs more attractive. Aside from practical considerations, such as differential cost and resources associated with PDs and EDs, there are certain populations for which paper rather than plastic may be the more appropriate and effective daily diary data entry portal. For example, although empirical evidence is equivocal,

several authors have voiced concerns regarding ED's susceptibility to being lost, damaged, or stolen in the school environment (cf. Savill-Smith & Kent, 2003). Previous computer experience, which is far less evident among older individuals, has been linked in several studies to the acceptability of EDs (e.g., Bernhardt et al., 2001). Moreover, the small screen of the PDA as an ED has been noted as a problem by numerous commentators (e.g., Embi, 2001), and others who work with older adults have noted that the high-pitched prompt of EDs may pose significant difficulties even for older individuals experiencing normal sensory aging (Patricia Parmelee, personal communication, July 6, 2003). Bolger et al. (2003) voiced concern that researchers may hesitate to use high-cost EDs in studies of poverty. Indeed, Weinhardt, Kalichman, Smyth, Carey, and Heron (2005) recently reported that among at-risk inner-city residents, more than 10% of the PDAs used as EDs were lost within 2 weeks, along with the data stored in them. Use of an ED in communities with high crime rates may increase the exposure of research participants to victimization, although cell phone EDs may reduce the risk. In each of these situations, PDs may have an advantage over EDs as long as the questions being investigated can be evaluated adequately in PD format.

### *Some Sorely Needed Data for Decision Making Concerning Paper Versus Plastic*

Many observers of the paper versus plastic debate are not so much interested in which is superior for their purposes as in the pragmatic question of "Given limited resources, what is the risk of relying on the generally more economical paper diaries?" Despite our reservations about some aspects of Green et al.'s (2006) equivalence study, we view the basic strategy of examining equivalence of the results obtained with the two methods to be important. We need to know more about the degree of imprecision introduced by reliance on PD when ED is clearly the superior strategy.

We also need more systematic data on how the demands of data collection affect which research participants are recruited and retained. For instance, which patients will decline participation or drop out of a clinical trial because of the intrusion of ED methods or because they cannot safely keep PDAs or other hardware in their home environments? This obviously speaks to the external validity of results of such trials, but it also speaks to the internal validity when drop out or lost data are considered treatment failure in intent-to-treat analyses.

### *Two Modest Proposals for Fair Comparisons of PDs and EDs*

A fair comparison of PD and ED data equivalence begins by acknowledging that within-day PD reports cannot be

verified reliably, and the comparison would therefore rely on a once-daily diary protocol in which PD diary entries are verified through postmarks, which admittedly allow participants to complete diaries the following day. The variables under study could include discrete behaviors or experiences (Bolger et al., 2003; Stone & Shiffman, 2002), recalled experiences of the day, and behaviors that are likely to be influenced by recalled experience (Redelmeier et al., 2003; Wirtz et al., 2003). This study would examine same-day relationships, cross-day lags, or weekday-weekend cycles (Tennen, Affleck, & Armeli, 2003). Such a study would provide a fair comparison by Cooper and Richardson's (1986) criterion while acknowledging the limits of PDs. Paper might make a strong showing in this comparison.

A fair comparison of ED compliance-enhancement features versus participants' collaboration-verification expectations would cross participants' beliefs about compliance and verification with two methodological levels—run of the mill versus state of the art—in a  $2 \times 2$  design. We imagine something akin to Broderick et al.'s (2003) ED compared with Green et al.'s (2006) Studies 2 and 3 EDs, crossed by instructions that highlight the importance of compliance and underscore verification potential versus instructions that mention neither compliance nor verification. In all four cells, however, data entry would be verifiable. If compliance-enhancement features carry the day, we would expect a main effect advantage on verified compliance and data quality for ED. However, if instructions, collaboration, and expectations regarding verification are key, we would expect an instructional set main effect on protocol adherence and data quality. ED advocates might predict a statistical interaction in which participants using the state-of-the-art ED and receiving collaboration-enhancing/verification instructions demonstrate the highest level of data quality. This sort of design, akin to a fully crossed placebo-controlled study, would be more evenhanded and informative than those offered by ED or PD adherents.

### Conclusions

Green et al.'s (2006) studies begin to provide much needed balance in efforts to examine the relative benefits and limitations of PD and ED methods. We have noted ways in which advocates of PDs and EDs have used methods that create unfair comparisons, we reviewed the relative strengths and limitations of PDs and EDs, and we offered study designs that we believe would offer more evenhanded comparisons. An old Russian proverb, “*doveray, no proveryay*,” translates roughly to “trust, but verify.” Thus far, many PD studies have been heavy on trust and light on verification. ED studies, on the other hand, have underscored verification. By combining Green et al.'s call for a collaborative relationship between investigators and diary participants with the admonitions of ED advocates to verify

diary entries, we can move beyond partisan debates and toward the application of methods that best examine the dynamic processes of daily life.

### References

- Affleck, G., Tennen, H., Zautra, A., Urrows, S., Abeles, M., & Karoly, P. (2001). Women's pursuit of personal goals in daily life with fibromyalgia: A value-expectancy analysis. *Journal of Consulting and Clinical Psychology, 69*, 587–596.
- Barrett, L. F. (1997). The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personality and Social Psychology Bulletin, 23*, 1100–1110.
- Barta, W. D., Kiene, S. M., Tennen, H., & Abu-Hasaballah, K. (2006). *Using diary methods to understand inconsistent condom use among profoundly disadvantaged individuals living with HIV: A research prospectus*. Unpublished manuscript.
- Bernhardt, J. M., Strecher, V. J., Bishop, K. R., Potts, P., Madison, E. M., & Thorp, J. (2001). Handheld computer-assisted self-interviews: User comfort level and preferences. *American Journal of Health Behavior, 25*, 557–563.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616.
- Broderick, J. E., Schwartz, J. E., Shiffman, S., Hufford, M. R., & Stone, A. A. (2003). Signaling does not adequately improve diary compliance. *Annals of Behavioral Medicine, 26*, 139–148.
- Carpenter, J. S., & Monahan, P. O. (2004). Accuracy of subjective hot flush reports compared with continuous sternal skin conductance monitoring. *Obstetrics & Gynecology, 104*, 1322–1326.
- Cohen, L. H., Gunthert, K. C., Butler, A. C., O'Neill, S. C., & Tolpin, L. H. (2005). Daily affective reactivity as a prospective predictor of depressive symptoms. *Journal of Personality, 73*, 1687–1714.
- Collins, R. L., Morsheimer, E. T., Shiffman, S., Paty, J. A., Gnys, M., & Papanonatos, G. (1998). Ecological momentary assessment in a behavioral drinking moderation training program. *Experimental and Clinical Psychopharmacology, 6*, 306–315.
- Cooper, W. H., & Richardson, A. J. (1986). Unfair comparisons. *Journal of Applied Psychology, 71*, 179–184.
- Dabbs, J. M., Jr., Bassett, J. F., & Dyomina, N. V. (2003). The Palm IAT: A portable version of the Implicit Association Test. *Behavior Research Methods, Instruments, and Computers, 35*, 90–95.
- DeLongis, A., Capreol, M. J., Holtzman, S., O'Brien, T. B., & Campbell, J. (2004). Social support and social strain among husbands and wives: A multi-level analysis. *Journal of Family Psychology, 18*, 470–479.
- Embi, P. J. (2001). Information at hand: Using handheld computers in medicine. *Cleveland Clinic Journal of Medicine, 68*, 840–853.
- Feinn, R., Tennen, H., Cramer, J., & Kranzler, H. (2003). Measurement and prediction of medication compliance in problem

- drinkers. *Alcoholism: Clinical and Experimental Research*, 27, 1286–1292.
- Gable, S. L., Reis, H. T., & Elliot, A. J. (2000). Behavioral activation and inhibition in everyday life. *Journal of Personality and Social Psychology*, 78, 1135–1149.
- Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*, 11, 87–105.
- Hufford, M. R., Shields, A. L., Shiffman, S., Paty, J., & Balabanis, M. (2002). Reactivity to ecological momentary assessment: An example using undergraduate problem drinkers. *Psychology of Addictive Behaviors*, 16, 205–211.
- Korotitsch, W. J., & Nelson-Gray, R. O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychological Assessment*, 11, 415–425.
- Levit, M., Huber, R., Batliner, A., & Noeth, E. (October, 2001). *Use of prosodic speech characteristics for automated detection of alcohol intoxication*. Paper presented at the Prosody in Speech Recognition and Understanding Conference, Red Bank, NJ.
- Litt, M. D., Cooney, N. L., & Morse, P. (1998). Ecological momentary assessment (EMA) with treated alcoholics: Methodological problems and potential solutions. *Health Psychology*, 17, 48–52.
- McKenzie, S., Paty, J., Grogan, D., Rosano, M., Curry, L., Sciarappa, K., & Hufford, M. (2004). Proving the eDiary Dividend. *Applied Clinical Trials*, 13, 54–68.
- Ptacek, J. T., Smith, R. E., Espe, K., & Raffety, B. (1994). Limited correspondence between daily coping reports and retrospective coping recall. *Psychological Assessment*, 6, 41–49.
- Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: A randomized trial. *Pain*, 104, 187–194.
- Rusting, C. L., & Larsen, R. J. (1998). Diurnal patterns of unpleasant mood: Associations with neuroticism, depression, and anxiety. *Journal of Personality*, 66, 85–113.
- Savill-Smith, C., & Kent, P. (2003). *The uses of palmtop computers for learning*. London: Learning and Skills Development Agency.
- Silvia, P. J. (2002). Self-awareness and emotional intensity. *Cognition and Emotion*, 16, 195–216.
- Stone, A., Kessler, R., & Haythornthwaite, J. (1991). Measuring daily events and experiences: Decisions for the researcher. *Journal of Personality*, 59, 575–608.
- Stone, A. A., Schwartz, J. E., Neale, J. M., Shiffman, S., Marco, C. A., Hickcox, M., et al. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology*, 74, 1670–1680.
- Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, 24, 236–243.
- Stone, A. A., Shiffman, S., & DeVries, M. W. (1999). Ecological momentary assessment. In D. Kahneman & E. Diener (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 26–39). New York: Russell Sage Foundation.
- Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials*, 24, 182–199.
- Swendsen, J., Tennen, H., Carney, M. A., Affleck, G., Willard, A., & Hromi, A. (2000). Mood and alcohol consumption: An experience sampling test of the self-medication hypothesis. *Journal of Abnormal Psychology*, 109, 198–204.
- Tennen, H., Affleck, G., & Armeli, S. (2003). Daily processes in health and illness. In J. Suls & K. Wallston (Eds.), *The social psychological foundations of health and illness* (pp. 495–529). Oxford, England: Blackwell.
- Tennen, H., Affleck, G., & Armeli, S. (2005). Personality and daily experience revisited. *Journal of Personality*, 73, 1465–1484.
- Tennen, H., Affleck, G., & Zautra, A. (in press). Depression history and coping with chronic pain: A daily process analysis. *Health Psychology*.
- Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology*, 59, 291–297.
- Tugade, M. M., Conner, T., & Barrett, L. F. (in press). Assessment of mood. In S. Ayers, A. Baum, C. McManus, S. Newman, K. Wallston, J. Weinman, & R. West (Eds.), *The Cambridge handbook of psychology, health, and medicine* (2nd ed.). Cambridge, England: Cambridge University Press.
- Weiler, J. M. (2005, April). *Comparison of controlled clinical trials of paper versus ePRO or ePRO versus ePRO*. Paper presented at the Drug Information Association Symposium, Arlington, VA.
- Weiler, K., Christ, A. M., Woodworth, G. G., Weiler, R. L., & Weiler, J. M. (2004). Quality of patient-reported outcome data captured using paper and interactive voice response diaries in an allergic rhinitis study: Is electronic data capture really better? *Annals of Allergy, Asthma & Immunology*, 92, 335–339.
- Weinhardt, L. S., Kalichman, S. C., Smyth, J. M., Carey, M. P., & Heron, K. (2005, April). *Ecological momentary assessment of sexual risk behavior among inner-city public sexually transmitted disease clinic patients I: Feasibility*. Poster session presented at the Society of Behavioral Medicine 26th Annual Meeting, Boston.
- Wirtz, D., Kruger, J., Scollon, N., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science*, 14, 520–524.

Received August 11, 2005

Revision received September 21, 2005

Accepted September 21, 2005 ■