Soc Indic Res (2010) 99:269–283
DOI 10.1007/s11205-010-9578-7

A Comparison of Affect Ratings Obtained with Ecological Momentary Assessment and the Day Reconstruction Method

Samantha Dockray · Nina Grant · Arthur A. Stone · Daniel Kahneman · Jane Wardle · Andrew Steptoe

Accepted: 9 January 2010 / Published online: 19 January 2010
© Springer Science+Business Media B.V. 2010

Abstract Measurement of affective states in everyday life is of fundamental importance in many types of quality of life, health, and psychological research. Ecological momentary assessment (EMA) is the recognized method of choice, but the respondent burden can be high. The day reconstruction method (DRM) was developed by Kahneman and colleagues (*Science*, 2004, *306*, 1776–1780) to assess affect, activities and time use in everyday life. We sought to validate DRM affect ratings by comparison with contemporaneous EMA ratings in a sample of 94 working women monitored over work and leisure days. Six EMA ratings of happiness, tiredness, stress, and anger/frustration were obtained over each 24 h period, and were compared with DRM ratings for the same hour, recorded retrospectively at the end of the day. Similar profiles of affect intensity were recorded with the two techniques. The between-person correlations adjusted for attenuation ranged from 0.58 (stress, working day) to 0.90 (happiness, leisure day). The strength of associations was not related to age, educational attainment, or depressed mood. We conclude that the DRM provides reasonably reliable estimates both of the intensity of affect and variations in affect over the day, so is a valuable instrument for the measurement of everyday experience in health and social research.

Keywords Affect · Measurement · Diurnal variation

S. Dockray · N. Grant · A. Steptoe (✉)
Psychobiology Group, Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, London WC1E 6BT, UK
e-mail: a.steptoe@ucl.ac.uk

A. A. Stone
Department of Psychiatry and Behavioral Science, Stony Brook University, Stony Brook, NY, USA

D. Kahneman
Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ, USA

J. Wardle
Health Behaviour Research Centre, Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, London WC1E 6BT, UK

1 Introduction

The assessment of mood and experience in everyday life is important in a number of domains. For example, research on mood disorders has noted distinctive patterns of mood variation over the day (Peeters et al. 2006; Stone et al. 1996), while clinical studies of pain highlight the relevance of associations between daily activities and pain experience (Krueger and Stone 2008). Recent economic valuations of non-market goods have emphasized the need to consider experienced utility and hedonic responses to experience, as well as decision utility (Dolan and Kahneman 2008). Studies relating affective state and biology have shown more consistent associations when affect is assessed repeatedly over the day compared with global retrospective reports (Bhattacharyya et al. 2008; Steptoe et al. 2007).

Ecological momentary assessment (EMA, also known as the experience sampling method) involves repeated sampling of experiences and mood in real time in the natural environment. It is considered the gold standard for the assessment of mood and experience in everyday life, since recall biases and heuristic biases are reduced, and ecological validity is maximized (Shiffman et al. 2008). However, the respondent burden with EMA is high, and assessments may be reactive. There are also circumstances in which collection of EMA data is difficult, as with some high stress occupational groups.

The day reconstruction method (DRM) was developed by Kahneman and colleagues as an alternative technique for assessing experiences and affect in everyday life (Kahneman et al. 2004). It builds on time-budget measurement and EMA, and involves the retrospective recall of the study period as a continuous sequence of episodes. Each episode is described in terms of onset and duration, location, social interaction, and activity. Participants also rate the episodes on a series of affect scales. Although the DRM involves recall, it is designed to increase the accuracy of emotional recollection by involving retrieval of the specifics of each episode (Robinson and Clore 2002). Using the DRM in a large sample of working women, Stone et al. (2006) found higher levels of positive affect in the evening, with a secondary peak in the morning. Ratings of depression, anger, and worry tended to diminish over the day, while tiredness increased. The DRM has been used in health studies to evaluate the experience of pain in a community sample (Krueger and Stone 2008), in studies of work and well-being (Kopperud and Vitterso 2008), personality and affect (Srivastava et al. 2008), behavior and well-being (White and Dolan 2009), and in clinical studies of depression and heart rate variability (Bhattacharyya et al. 2008).

Direct comparisons between EMA and DRM assessments of affect in the same individual over the same time period have not yet been published. In their original description of the method, Kahneman et al. (2004) showed that the profiles over the day of affect measured with the DRM closely resembled those of a previous EMA Study (Stone et al. 1998), but did not provide a direct comparison. More recently, Krueger and Schkade (2008) found that the test–retest reliability of DRM measures aggregated over the day averaged 0.45–0.65 for different affect ratings from the same individuals tested 2 weeks apart.

The present analysis was therefore designed to assess the strength of associations between EMA and DRM assessments of affect in the same individuals over the same time period. Each EMA measure was compared with affect assessed with the DRM over the corresponding hour of the day. Measurements were obtained from a sample of working women on work days and leisure days, so that that the robustness of associations on different conditions could be analyzed. We investigate three questions. First, whether the profiles and the intensity of affect over the work and leisure days were similar with the two techniques. Second, whether the between-person correlations between techniques was sufficiently high to provide validation of the DRM. Third, whether there was systematic variation in the strength of associations between EMA and DRM affect measures, depending on factors such as age, socioeconomic status (SES) as defined by educational attainment, or depressed mood.

2 Method

2.1 Participants

Participants in this study were 94 women aged 21–54 years working at University College London. They were recruited via direct email contact and flyers posted in communal work areas. Exclusion factors were pregnancy, serious illness or disease and use of medications, including psychotropics and non-steroidal anti-inflammatories. The study was approved by the University College London Research Ethics Committee, and all participants gave written consent.

2.2 Design

Each participant completed two 24 h periods of EMA and DRM assessment beginning in the later afternoon, one on a working day, and the other starting after work on a Friday and continuing until Saturday afternoon (leisure day). The order of assessment periods was counterbalanced across participants, so that half began with the working day and half with the leisure day. During each period, 6 sets of EMA ratings were obtained, two over the evening and four on the next day. The DRM was completed online at the end of each 24 h period. This design differs from that used in earlier DRM studies in that the recall period was split by the night.

2.3 EMA Assessments

EMA assessments were made in conjunction with saliva sampling for cortisol (to be described elsewhere). Participants were given a sampling diary and were instructed to collect saliva samples and complete the EMA assessments at 6 time points: while in the research office (typically 17:00–18:00 pm), at bedtime, 30 min after waking the next morning, 10:00 h, 12:00 noon, and at 15:00 h. No external aids were provided to ensure sampling at the requested time points. Participants were asked to rate their current feelings on a series of 5-point scales were 1 = *not at all*, and 5 = *very much*. The affects analyzed in this report were happy, tired, stressed and frustrated or angry.

2.4 DRM Assessments

The DRM was completed online at the end of each 24 h period. Participants were asked to recall the monitoring period as a continuous series of episodes, like a movie. Each episode was defined in terms of time of onset and duration, location, social situation and activity. After the complete 24 h had been reconstructed, participants rated their feelings during each episode on a series of 7-point scales from 0 = *not at all* to 6 = *very much*. The affective states analyzed here were happy, tired, worry, feeling hassled, angry and frustrated. Several other affective states were rated in the DRM that are not analyzed here because they did not correspond directly to EMA assessments (namely feeling impatient, criticized, competent, warm and friendly, and enjoyment).

2.5 Other Measures

Depressed mood was assessed using the Center for Epidemiologic Studies Depression scale (CES-D), a well established questionnaire measure (Radloff 1977). The standard cut-off of 16 was used to categorize at high risk of depressive disorder. Participants' marital status was assessed by questionnaire, and the sample was divided into those with and without a marital relationship. Educational qualifications were measured, and participants were divided according to whether they had a college degree. Income was classified into eight categories and subsequently divided into three bands: <£25,000, £25,000–£34,999, and >£35000. Participants were also questioned about their average work hours per week.

2.6 Procedure

Participants attended the research office individually at the end of a working day. Half the participants were scheduled for Monday to Thursday, so the next 24 h would constitute the work day, while the remainder attended on a Friday, immediately before the leisure day assessment. The procedure was explained, signed consent obtained, and participants were instrumented for 24 h measurements of heart rate and heart rate variability (not described here). At the end of the assessment period, participants returned to the research office to hand over the EMA diaries and complete the DRM online either at their workplace or home on work days. On leisure days, the DRM was completed online at home, and the materials were returned to the study team on the following Monday. The two assessment days were separated by a minimum of 2 days, and the second assessment day was scheduled within a maximum of 14 days of the first. Participants received a small honorarium for their involvement.

2.7 Statistical Analysis

The DRM provides a continuous sequence of activities with associated affect ratings. However, the episodes vary widely in number, timing and duration. One individual might be carrying out the same activity for a 2 h period, so have a single set of affect ratings, while another might have reported three different activities, each with associated affect ratings. In order to relate affect derived from the DRM to EMA measures, we derived measures for every hour of the monitoring period. Over every hour, the corresponding DRM episode was identified for each individual. If the hour involved more than one DRM episode, then a weighted average of affect ratings was computed. We then identified the hour in which each EMA rating was taken, and matched this with the appropriate hour of DRM affect. For example, if the 10 am EMA assessment was conducted at 10:23 am, we matched this with the average DRM rating for 10:00–11:00 am. In the case of the bedtime EMA ratings, comparison was made with DRM ratings for the previous hour. The result was six pairs of ratings of EMA and DRM affect (17:00 h, bedtime, waking + 30 min, 10:00, 12:00, and 15:00 h) for both work and leisure days for each affect.

There were two affect measures for which exactly the same wording was used in the EMA and DRM methods: happiness and tiredness. The other two EMA affects (frustrated/

angry and stressed) had no direct correspondence in the DRM. Accordingly, we associated EMA frustrated/angry with both angry and frustration ratings on the DRM. The EMA stress ratings were correlated with DRM ratings of worry and hassled. EMA ratings were obtained on a 1–5 scale, while DRM ratings ranged from 0 to 6. In order to make direct comparisons, each was scaled to the same metric ranging from 0 to 24, where 0 represented the lowest and 24 the most intense affect.

The first set of analyses assessed profiles of affect across the six ratings for each day, and whether the mean intensity of affect differed for each pair of EMA/DRM measures across time. Multilevel modelling with MLwiN software (Rasbash et al. 2001) was used to carry out repeated measures multivariate analysis. Polynomial contrasts were modeled so as to capture the pattern of affect by measure type across the day. Separate models were tested for work and leisure days involving each pair (for happy and tired) or trio (for stressed and angry/frustrated) of EMA/DRM affect ratings. A hierarchical two level structure was employed, with measurement occasion modelled at level 1, nested within individuals at level 2. In addition to attending to the correlated structure of the data within individuals, this approach also allowed participants with incomplete data to be included. The significance of estimated coefficients was tested using the Wald criterion, and contrasts between coefficients were tested using χ^2 (Raudenbush and Bryk 2002) with Bonferroni adjustments for multiple comparisons. When significant interactions between measure and time were identified, effects were analyzed post hoc using Tukey's least significant difference (LSD) test.

The second set of analyses assessed the between-person associations between the EMA and DRM measures using product-moment correlations. Twelve correlations (six time points on the 2 days) were calculated for each pair of affect ratings. These were subsequently aggregated by converting the correlation coefficients into z scores, averaging the z scores, and then converting them back to correlations (r). The result was an average correlation for every pair of ERM/DRM affect ratings for the working and leisure days.

Third, we investigated whether variations in the strength of correlations between ERM and DRM ratings were systematically associated with the background characteristics of the participants. Three factors were analyzed: age, educational attainment, and depressed mood. We compared the strength of aggregate correlations using Fisher's Z between participants older or younger than 30 years, those with and without a college degree, and individuals with CES-D scores above and below the cut-off of 16. Analyses were carried out using SPSS v.14, MLwin v2.02 and MedCalc.

3 Results

The characteristics of participants are summarized in Table 1. Participants were aged 32 years on average, and the majority were of white European origin. Fewer than half had college degrees, while 57% were married. Participants reported working an average of 41.1 h per week, of which 38.9 h was in the workplace. The proportion of respondents with CESD scores $\geq$16 was 28.4%.

3.1 Data Availability

Table 2 summarizes the data included in these comparisons. It can be seen that 93 of the 94 participants completed the DRM for the work day, but only 79 on the leisure day. The main

Table 1 Participant characteristics

	Mean ± SD or *N* (%)
Age (years)	32.6 ± 8.4
Marital status (married/cohabiting)	54 (57.4%)
Ethnicity (white European)	75 (78.9%)
Education (college degree)	40 (44.4%)
Income	
<£25,000	35 (37.2%)
£25,000–34,999	43 (45.7%)
>£35,000	16 (17.0%)
CES-D depression score	12.5 ± 9.3
Working hours on site	38.9 ± 6.4
Total working hours	41.1 ± 8.3

Table 2 Data available for EMA/DRM comparisons

Measure	*N* participants (%)	Number of datapoints available[a] Mean (SD)
Work day		
Day reconstruction method	93 (98.9%)	5.32 (1.21)
Ecological momentary assessment	88 (93.6%)	5.31 (1.21)
Both	86 (91.5%)	
Leisure day		
Day reconstruction method	79 (84.0%)	5.41 (1.15)
Ecological momentary assessment	91 (96.8%)	5.31 (1.21)
Both	74 (78.7%)	

[a] Range 1–6

reason for data loss on the leisure day was failure in accessing the DRM website. A small proportion of participants (<10%) did not have EMA assessments on both days, due to the loss or non-return of the EMA diaries by the participant. Thus data for the EMA/DRM comparisons were available from 86 participants on the working day, and 74 on the leisure day. Comparisons between individuals with and without data on the working day found no differences in terms of demographic characteristics. However, women who did not provide leisure day data were significantly younger than the remainder ($t = 2.07$, $p = 0.041$). Within each day, the average number of data points available ranged from 5.26 to 5.41 (out of the possible range of 1–6), and did not differ between the two techniques. Most data was lost at 15:00 h (the last sampling point), with only 58 (67%) of participants on the work day and 54 (73%) on the leisure day providing data at this point. The reported times of EMA assessment averaged 17:26 h ± 35 m, 23:03 ± 53 m, 7:18 h ± 52 m, 10:05 h ± 16 m, 12:02 h ± 12 m and 15:06 h ± 27 m for the work day, and 17:30 h ± 36 m, 23:32 h ± 66 m, 8:26 h ± 72 m, 10:16 ± 35 m, 12:15 h ± 42 m, and 15:11 h ± 35 m for the leisure day.

3.2 Profile of Affect Over the Day

The profile of happiness and tiredness measured with the two techniques on work and leisure days is summarized in Fig. 1. It can be seen that happiness was low in the early part of the work day, while being maintained at a higher level on the leisure day. Repeated multivariate analyses of happiness ratings showed a main effect of time on both the work and leisure day but no effect of measure (EMA vs. DRM). Polynomial contrasts indicated a significant cubic change on the work day and a linear effect on the leisure day. Thus the absolute ratings of happiness with EMA and DRM techniques were indistinguishable, and the two methods showed the same pattern of change over the assessment periods. Tukey's LSD tests confirmed that ratings were lower at Wake + 30 than all other time points on the work day using both methods ($p < 0.0.05$). On the leisure day, significant differences between 17:00 h and 15:00 h were reported for both methods.

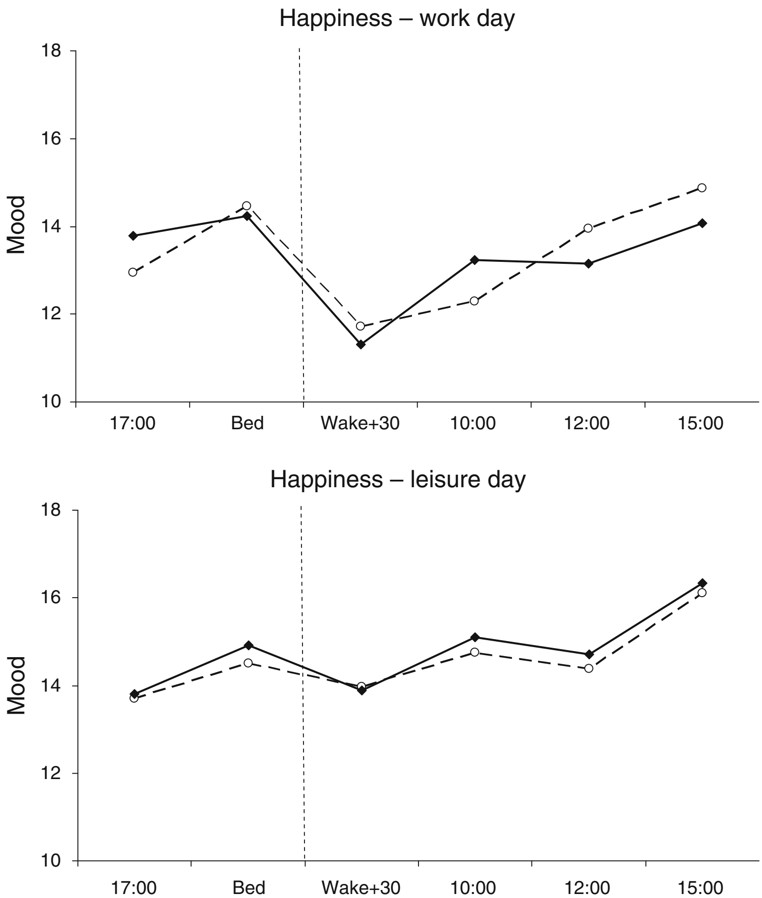

Fig. 1 Mean ratings of happiness on the work (*upper panel*) and leisure (*lower panel*) days. Samples were obtained at 17:00 h, bed-time, 30 min after waking (Wake + 30), 10:00 am, 12:00 noon, and 15:00 h. *Solid line/diamond*: EMA happiness; *dashed line/circle*: DRM happiness. The *vertical dashed line* indicates the night separating the two monitoring periods of each day

The multilevel analyses of tiredness indicated a similar pattern across work and leisure days, with tiredness at its greatest at bed-time on both days (Fig. 2). On the work day there was no main effect of measure type, but a quadratic effect of time ($\chi^2 = 6.02$, $df = 2$, $p < 0.03$). Participants were more tired at bedtime than all other time points. Additionally, tiredness ratings at Wake + 30 were higher than those later in the day ($p < 0.05$). On the leisure day, there was a significant interaction between time and measure ($\chi^2 = 9.23$, $df = 1$, $p < 0.02$). Post hoc tests indicated that the EMA ratings taken at bedtime were higher than the DRM ratings, and were also greater at 10:00 and 15:00 h ($p < 0.05$). DRM ratings on the leisure day did not increase between 17:00 h and bedtime.

The profile of EMA stress ratings is summarized in Fig. 3, together with DRM ratings of worry and feeling hassled. Multilevel modelling found no significant effects of measure type either in the analyses of EMA stress and DRM worry, or in the trio of EMA stress, DRM hassle and DRM worry on either work or leisure day. On the work

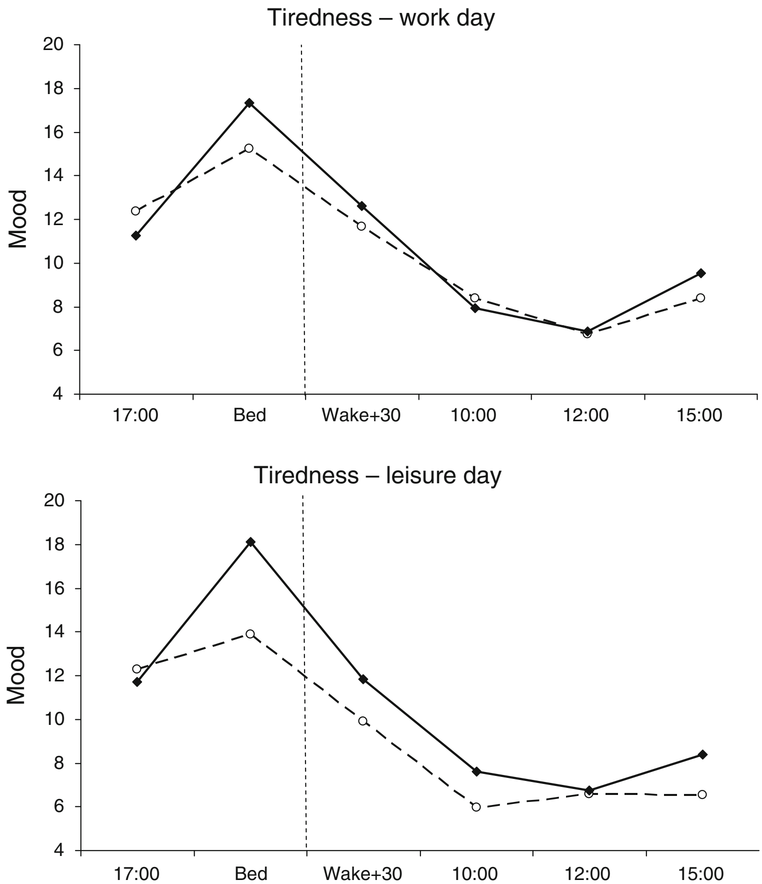

Fig. 2 Mean ratings of tiredness on the work (*upper panel*) and leisure (*lower panel*) days. Samples were obtained at 17:00 h, bed-time, 30 min after waking (Wake + 30), 10:00 am, 12:00 noon, and 15:00 h. *Solid line/diamond*: EMA tiredness; *dashed line/circle*: DRM tiredness. The *vertical dashed line* indicates the night separating the two monitoring periods of each day

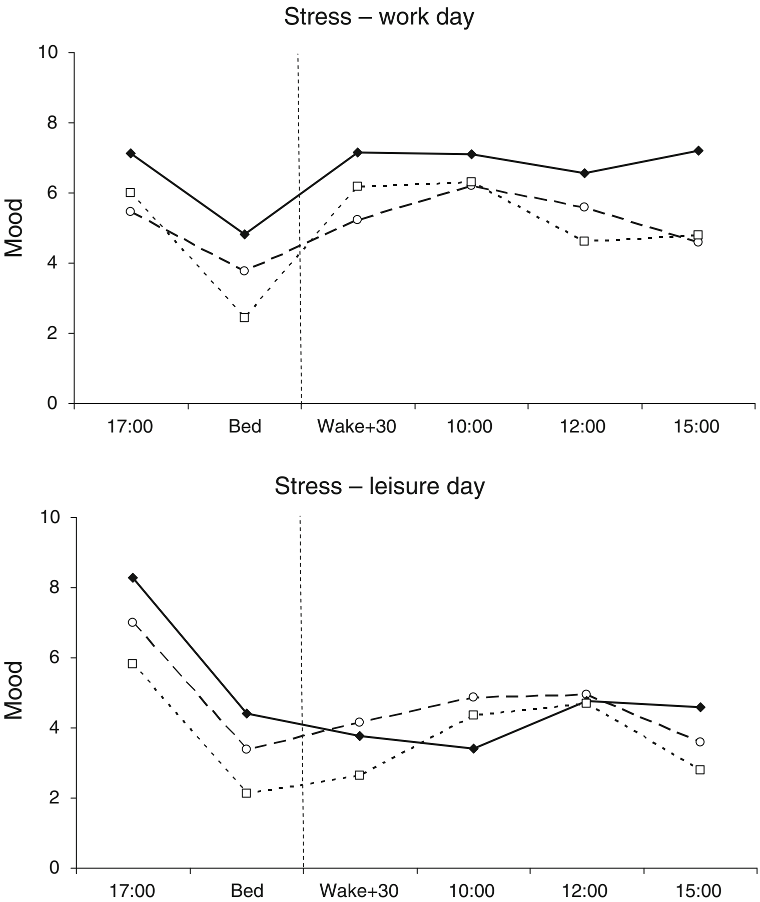

Fig. 3 Mean ratings of stress on the work (*upper panel*) and leisure (*lower panel*) days. Samples were obtained at 17:00 h, bed-time, 30 min after waking (Wake + 30), 10:00 am, 12:00 noon, and 15:00 h. *Solid line/diamond*: EMA stress; *dashed line/circle*: DRM worry; *dotted line/square*: DRM hassled. The *vertical dashed line* indicates the night separating the two monitoring periods of each day

day, EMA stress ratings and DRM worry and hassle ratings were lower at bedtime than all other samples, while on the leisure day, ratings were higher at 17:00 h than all other times ($p < 0.05$).

EMA assessments of anger/frustration are depicted in Fig. 4, together with DRM anger and frustration. The multilevel repeated measures multivariate analyses including the EMA and both DRM measures indicated a significant measure by time effect on both the work ($\chi^2 = 13.23$, $df = 1$, $p < 0.02$) and leisure day ($\chi^2 = 9.46$, $df = 1$, $p < 0.02$). On the work day, the DRM anger ratings were lower than EMA anger/frustration ratings at all time points, while DRM frustration ratings were greater than EMA at 17:00 h and 10:00 h ($p < 0.05$). A similar pattern was observed on the leisure day, with DRM anger ratings being lower at 17:00 h, bed-time, Wake + 30 and 15.00 than corresponding EMA ratings, with DRM frustration being higher at 17:00 h.

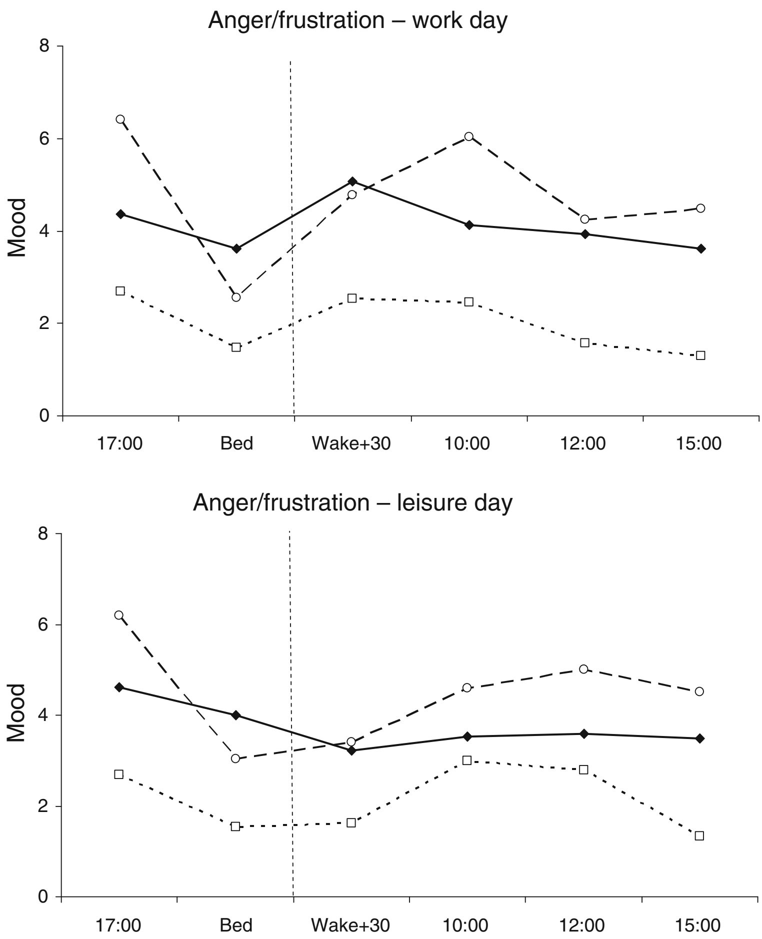

Fig. 4 Mean ratings of anger/frustration on the work (*upper panel*) and leisure (*lower panel*) days. Samples were obtained at 17:00 h, bed-time, 30 min after waking (Wake + 30), 10:00 am, 12:00 noon, and 15:00 h. *Solid line/diamond*: EMA anger/frustration; *dashed line/circle*: DRM frustration; *dotted line/square*: DRM anger. The *vertical dashed line* indicates the night separating the two monitoring periods of each day

3.3 Correlations Between EMA and DRM Ratings

The correlations between EMA and DRM ratings are detailed in Table 3. Both the correlations between average values and the aggregate correlations derived using Fisher's z were highly significant ($p < 0.001$). As might be expected the correlations between averaged values were stronger than those in individual time point comparisons. The mean correlations aggregated from individual time points ranged from 0.61 (EMA stress vs. DRM worried on the leisure day), to 0.35 (EMA stress vs. DRM hassled on the work day). There were no systematic variations in the strength of correlations across the different moods or assessment days. The range of correlations for individual time points is also shown for each comparison in Table 3. All individual time point correlations were significant at $p < 0.01$ except for three: EMA angry-frustrated vs. DRM anger at 3:00 pm on the work day ($r = 0.28$) and leisure days ($r = 0.18$), and EMA stressed vs DRM hassled at

Table 3 Associations between EMA and DRM mood ratings

EMA mood/ DRM mood	Work day			Leisure day		
	Mean rating *r*	Mean of individual time point comparisons *r*	Range	Mean rating *r*	Mean of individual time point comparisons *r*	Range
Happy/happy	0.72	0.52	0.31–0.57	0.61	0.46	0.34–0.63
Tired/tired	0.63	0.53	0.34–0.65	0.67	0.54	0.32–0.67
Angry-frustrated/ angry	0.60	0.45	0.28–0.59	0.64	0.46	0.18–0.60
Angry-frustrated/ frustrated	0.62	0.43	0.29–0.54	0.72	0.57	0.50–0.66
Stressed/worried	0.58	0.41	0.33–0.48	0.79	0.61	0.47–0.70
Stressed/hassled	0.52	0.35	0.20–0.43	0.59	0.45	0.31–0.65

All $p < 0.001$

noon on the work day ($r = 0.20$). There were no differences in the strength of associations between participants with complete and partial EMA data.

3.4 Adjustment of Correlations for Attenuation

The strength of the association between the two methods is attenuated in proportion to the degree of error in each method. We therefore used the method recommended by Nunnally and Bernstein (1994), as detailed in Krueger and Schkade (2008) to estimate attenuation of the true association by measurement error. We applied the formula $\rho_{xy} = r_{xy}/\sqrt{(r_{xx} \times r_{yy})}$, where ρ_{xy} = the true correlation between x and y; r_{xy} = observed correlation between x and y; r_{xx} = reliability of x; r_{yy} = reliability of y. We do not have precise estimates of intra-measure reliability. However, inspection of Figs. 1, 2, 3, 4 indicates that the pattern of affect was relatively stable between 10:00 am and 12:00 noon. We therefore used the within-measure correlation between these points to estimate r_{xx} and r_{yy}. Separate analyses were conducted for work and leisure days.

These analyses are summarized in Table 4. It is evident that adjustment for attenuation resulted in substantial increases in the strength of correlations between EMA and DRM ratings. For example, tiredness correlations rose from 0.53 and 0.54 for work and leisure days to 0.75 and 0.86, respectively. The greatest adjusted correlations were for happiness and the relationship between EMA stress and DRM worry on the leisure day (0.90). The lowest values (0.58) were for EMA stress and worry and feeling hassled on the work day.

3.5 Moderators of EMA/DRM Associations

The comparison of strength of correlations between EMA and DRM ratings in relation to age, educational status and depressed mood uncovered only two significant differences: the correlation between happiness ratings on the working day between the two methods was lower in depressed than nondepressed individuals ($r = 0.39$ and 0.76, $z = 2.18$, $p = 0.029$), and the association between EMA stress and DRM hassle ratings on the leisure day was significantly greater in older than younger participants ($r = 0.74$ and 0.38, $z = 2.18$, $p = 0.029$). However, in view of the fact that 36 comparisons were made across

Table 4 Correlations between EMA and DRM affect, adjusted for attenuation

EMA mood/DRM mood	r_{xy}	r_{xx}	r_{yy}	ρ_{xy} disattenuated correlation
Work day				
Happy/happy	0.52	0.71	0.76	0.71
Tired/tired	0.53	0.66	0.75	0.75
Angry-frustrated/angry	0.45	0.49	0.71	0.76
Angry-frustrated/frustrated	0.43	0.49	0.62	0.78
Stressed/worried	0.41	0.58	0.87	0.58
Stressed/hassled	0.35	0.58	0.62	0.58
Leisure day				
Happy/happy	0.46	0.42	0.62	0.90
Tired/tired	0.54	0.55	0.72	0.86
Angry-frustrated/angry	0.46	0.62	0.93	0.61
Angry-frustrated/frustrated	0.57	0.62	0.85	0.79
Stressed/worried	0.61	0.51	0.51	0.90
Stressed/hassled	0.45	0.51	0.51	0.67

the different affect measures and days, a Bonferroni correction was applied, and these findings are no longer significant.

4 Discussion

This study was designed to assess the comparability of affect ratings obtained in real time using EMA techniques, and retrospectively with the DRM. The results indicated that the two methods produced very similar profiles of change over both work and leisure days. The average intensity of ratings was identical with the two methods for happiness and tiredness at almost all time points, whereas for the other comparisons (EMA stress vs. DRM worry and feeling hassled, and EMA angry/frustrated vs DRM anger and frustration), there were greater differences in intensity, probably due to differences in wording. The correlations between methods tended to be greater when the analyses were based on values averaged over the monitoring period, than when individual time points were compared. This suggests that the DRM may be useful in providing estimates of overall affect level over the day. The between-person correlations between methods were high after adjusting for attenuation, ranging from 0.58 to 0.90. Of the three factors tested as possible moderators of associations between measures, educational attainment had no impact, while there were inconsistent and possibly spurious effects of depressed mood and age.

When contrasting the affect ratings on the 2 days, it should be noted that the evening ratings for the leisure days were obtained on Friday evening, at the end of a working day when participants were anticipating the weekend. This arrangement was necessitated by equipment requirements of biological monitoring that will be described elsewhere. Consequently, there were rather few differences in affect between the first two time points (17:00 h and bedtime) on the 2 days.

Ratings of happiness showed the anticipated increase across the day and evening (Fig. 1). This effect was more marked on the work than leisure day, since on leisure days the participants did not exhibit especially low happiness ratings in the morning. We were not able to document the bimodal pattern of positive affect observed by Stone et al. (2006), with a decline in the early afternoon. This pattern may only emerge with more refined temporal analysis than was possible in this study. The convergence between the two methods of assessing affect was striking in the happiness analysis, with close comparability in intensity as well as pattern over the day. The mean correlation between the methods was 0.52 for the work and 0.46 for the leisure day, but when these figures were adjusted for attenuation, the associations increased to 0.71 and 0.90, respectively. These findings indicate that retrospective assessments of positive affect with the DRM are closely comparable with those obtained using EMA.

The correspondence between the methods of assessing tiredness was also good, with correlations adjusted for attenuation of 0.75 and 0.86 for work and leisure days. As in previous research, participants were more tired soon after waking up than in the middle of the day (Stone et al. 1996). Ratings of tiredness increased through the evening as might be expected. However, the DRM underestimated tiredness at bedtime, with a significant difference between measurement techniques on the leisure day (Fig. 2). The explanation for this discrepancy may be that EMA ratings were based on current affect, whereas the DRM was analyzed for the previous hour. Participants may not have been as tired throughout this period as when they finally went to bed.

We did not anticipate as close associations between the two techniques for the remaining comparisons involving stress and anger/frustration, since the EMA and DRM assessments were not identical. Feeling hassled or worried may be distinct from the experience of stress, while anger is likely to be a less common experience than the combination of anger and frustration. Ratings of stress on the work day were sustained at a relatively high level throughout, decreasing only at bedtime (Fig. 3). By contrast, levels of stress were lower on the leisure day. There was little diurnal variation in either EMA stress or the comparable DRM ratings of worry or feeling hassled, except for the drop at bed time. Ratings of anger and frustration were low through out the study and showed little systematic variation over time. Multi-level modeling revealed no differences between the techniques for stress, worry and feeling hassled, but DRM anger ratings were significantly lower than EMA anger/frustration ratings throughout the procedure, while DRM frustration ratings were higher during the day. Nevertheless, the correlation adjusted for attenuation range from 0.58 to 0.90 for these measures, suggesting that the DRM does a reasonably good job in evaluating these negative affective states.

We tested whether 3 factors could potentially moderate the strength of associations between EMA and DRM assessments. Age was considered since there may be age-related changes in the balance of positive and negative emotion that could lead to differential responding (Ross and Mirowsky 2008), and because age-related differences in the profile of affect over the day have been reported (Stone et al. 2006). We evaluated educational attainment in order to address the possibility that the complexity of the DRM tasks might tax some individuals, leading to poorer performance. Depression was tested because recall can be biased by prevailing mood, raising the possibility that more depressed individuals will show a tendency to recall the day more negatively, leading to a reduction in the capacity of the retrospective DRM to capture ongoing affective states. However, none of these factors was systematically related to this strength of EMA/DRM associations, suggesting that the DRM is robust in these respects.

The raw correlations between EMA and DRM methods were markedly increased when we corrected for attenuation due to measurement error within each technique. The application of this procedure should be interpreted cautiously, since we did not have pairs of temporally close estimates within each method that would provide precise within-method reliability coefficients. In order to estimate within-method reliability, we use the correlations between measures taken at two adjacent time points (10.00 h and 12:00 noon). These points were selected because they had the smallest time interval in the study, and because the average affect profiles showed relatively stable level across this period. However, the within-measure correlations will also have been influenced by genuine changes in affect between two time points, so the estimate of reliability of each measure might have increased if ratings closer in time had been available.

This study has a number of limitations. First, EMA ratings were collected using paper diary rather than electronic assessments, so the timing of data collection could not be confirmed objectively. Many EMA studies are conducted using electronic devices in which the timing of assessments is recorded, although there is some debate about whether electronic methods do provide more accurate data than pencil and paper diaries (Green et al. 2006; Tennen et al. 2006; Shiffman et al. 2008). This method may have resulted in some inaccuracy in data collection that could either have inflated or attenuated the true correlations. On average, participants reported completing EMA ratings around the requested time, and we used each individual's reported timing to select the DRM ratings included in the comparison. Theoretically it is possible that participants completed EMA ratings at different times from those reported. We are doubtful about this explanation, since the cortisol samples that were collected at the same time as EMA ratings (not discussed here), showed appropriate diurnal variation. Additionally, there were no differences in the strength of correlations between individuals with complete compared with partly complete EMA diaries; it seems implausible that those with incomplete EMA data filled in the measures simultaneously in a retrospective fashion, as might have been the case for participants with complete data. Second, the DRM ratings used in the comparisons were aggregated across the hour in which the EMA ratings obtained, and a more exact time frame for the DRM might have increased the strength of associations. Third, data were collected from a sample of working women in a major city, and results may not generalize to other populations. Fourth, it is possible that completion of the EMA increased participants' consciousness of their activities over the day, enabling them to complete the DRM more accurately than if they had been asked for this information without warning. Finally, this study was limited by the absence of directly comparable wording in the DRM for EMA ratings of stress and anger/frustration and by the different scales used for the two methods, and these are likely to have reduced the magnitude of correlations between methods.

Nevertheless, the findings suggest that the DRM does a reasonable job in capturing the profile of affect experienced in real-time as documented by the EMA, and that individual differences in affect assessed with the two methods are comparable. These findings complement results recently reported by Krueger and Schkade (2008), and indicate that the DRM is a valuable instrument for the measurement of everyday experience and activity. The DRM may prove a useful tool in social and health research when investigators seek to study associations between affective states in everyday life and behavior or biological processes. It may be particularly relevant when EMA measures are sometimes difficult to obtain (e.g. surgeons, emergency services), or when there are doubts about the reliability of real-time momentary measures.

Acknowledgments This research was supported by the National Institute on Aging, USA (AG13196), the Economic and Social Research Council, UK (Res-177-25-0005), and the Medical Research Council, UK. Jane Wardle is supported by Cancer Research UK and Andrew Steptoe by the British Heart Foundation.

References

Bhattacharyya, M. R., Whitehead, D. L., Rakhit, R., & Steptoe, A. (2008). Depressed mood, positive affect, and heart rate variability in patients with suspected coronary artery disease. *Psychosomatic Medicine, 70*, 1020–1027.

Dolan, P., & Kahneman, D. (2008). Interpretations of utility and their implications for the valuation of health. *Economics Journal, 118*, 215–234.

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.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science, 306*, 1776–1780.

Kopperud, K. H., & Vitterso, J. (2008). Distinctions between hedonic and eudaimonic well-being: Results from a day reconstruction study among Norwegian jobholders. *Journal of Positive Psychology, 3*, 174–181.

Krueger, A., & Schkade, D. (2008). The reliability of subjective well-being measures. *Journal of Public Economics, 92*, 1833–1845.

Krueger, A. B., & Stone, A. A. (2008). Assessment of pain: A community-based diary survey in the USA. *Lancet, 371*, 1519–1525.

Peeters, F., Berkhof, J., Delespaul, P., Rottenberg, J., & Nicolson, N. A. (2006). Diurnal mood variation in major depressive disorder. *Emotion, 6*, 383–391.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psycholological Measurement, 1*, 385–401.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128*, 934–960.

Ross, C. E., & Mirowsky, J. (2008). Age and the balance of emotions. *Social Science and Medicine, 66*, 2391–2400.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32.

Srivastava, S., Angelo, K. M., & Vallereux, S. R. (2008). Extraversion and positive affect: A day reconstruction study of person-environment transactions. *Journal of Research in Personality, 42*, 1613–1618.

Steptoe, A., Gibson, E. L., Hamer, M., & Wardle, J. (2007). Neuroendocrine and cardiovascular correlates of positive affect measured by ecological momentary assessment and by questionnaire. *Psychoneuroendocrinology, 32*, 56–64.

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., Schwartz, J. E., Schkade, D., Schwarz, N., Krueger, A., & Kahneman, D. (2006). A population approach to the study of emotion: Diurnal rhythms of a working day examined with the day reconstruction method. *Emotion, 6*, 139–149.

Stone, A. A., Smyth, J. M., Pickering, T., & Schwartz, J. (1996). Daily mood variability: Form of diurnal patterns and determinants of diurnal patterns. *Journal of Applied Social Psychology, 26*, 1286–1305.

Tennen, H., Affleck, G., Coyne, J. C., Larsen, R. J., & DeLongis, A. (2006). Paper and plastic in daily diary research: Comment on Green, Rafaeli, Bolger, Shrout, and Reis (2006). *Psychological Methods, 11*, 112–118.

White, M. P., & Dolan, P. (2009). Accounting for the richness of daily activities. *Psychological Science, 20*, 1000–1008.