

Original Paper

# Comparability of Emotion Dynamics Derived From Ecological Momentary Assessments, Daily Diaries, and the Day Reconstruction Method: Observational Study

Stefan Schneider, PhD; Doerte U Junghaenel, PhD; Tania Gutsche, BA; Hio Wa Mak, PhD; Arthur A Stone, PhD  
University of Southern California, Los Angeles, CA, United States

**Corresponding Author:**

Stefan Schneider, PhD  
University of Southern California  
Dornsife Center for Self-Report Science  
635 Downey Way  
Los Angeles, CA, 90089-3332  
United States  
Phone: 1 2138211774  
Email: [schneids@usc.edu](mailto:schneids@usc.edu)

## Abstract

**Background:** Interest in the measurement of the temporal dynamics of people's emotional lives has risen substantially in psychological and medical research. Emotions fluctuate and change over time, and measuring the ebb and flow of people's affective experiences promises enhanced insights into people's health and functioning. Researchers have used a variety of intensive longitudinal assessment (ILA) methods to create measures of emotion dynamics, including ecological momentary assessments (EMAs), end-of-day (EOD) diaries, and the day reconstruction method (DRM). To date, it is unclear whether they can be used interchangeably or whether ostensibly similar emotion dynamics captured by the methods differ in meaningful ways.

**Objective:** This study aims to examine the extent to which different ILA methods yield comparable measures of intraindividual emotion dynamics.

**Methods:** Data from 90 participants aged 50 years or older were collected in a probability-based internet panel, the Understanding America Study, and analyzed. Participants provided positive and negative affect ratings using 3 ILA methods: (1) smartphone-based EMA, administered 6 times per day over 1 week, (2) web-based EOD diaries, administered daily over the same week, and (3) web-based DRM, administered once during that week. We calculated 11 measures of emotion dynamics (addressing mean levels, variability, instability, and inertia separately for positive and negative affect, as well as emotion network density, mixed emotions, and emotional dialecticism) from each ILA method. The analyses examined mean differences and correlations of scores addressing the same emotion dynamic across the ILA methods. We also compared the patterns of intercorrelations among the emotion dynamics and their relationships with health outcomes (general health, pain, and fatigue) across ILA methods.

**Results:** Emotion dynamics derived from EMAs and EOD diaries demonstrated moderate-to-high correspondence for measures of mean emotion levels ( $\rho \geq 0.95$ ), variability ( $\rho \geq 0.68$ ), instability ( $\rho \geq 0.51$ ), mixed emotions ( $\rho = 0.92$ ), and emotional dialecticism ( $\rho = 0.57$ ), and low correspondence for measures of inertia ( $\rho \geq 0.17$ ) and emotion network density ( $\rho = 0.36$ ). DRM-derived measures showed correlations with EMAs and EOD diaries that were high for mean emotion levels and mixed emotions ( $\rho \geq 0.74$ ), moderate for variability ( $\rho = 0.38-0.54$ ), and low to moderate for other measures ( $\rho = 0.03-0.41$ ). Intercorrelations among the emotion dynamics showed high convergence across EMAs and EOD diaries, and moderate convergence between the DRM and EMAs as well as EOD diaries. Emotion dynamics from all 3 ILA methods produced very similar patterns of relationships with health outcomes.

**Conclusions:** EMAs and EOD diaries provide corresponding information about individual differences in various emotion dynamics, whereas the DRM provides corresponding information about emotion levels and (to a lesser extent) variability, but not about more complex emotion dynamics. Our results caution researchers against viewing these ILA methods as universally interchangeable.

(*J Med Internet Res* 2020;22(9):e19201) doi: [10.2196/19201](https://doi.org/10.2196/19201)

**KEYWORDS**

ecological momentary assessment; daily diaries; day reconstruction method; emotion dynamics; emotion variability; mobile phone

## Introduction

The use of intensive longitudinal assessments (ILAs) in medical research has risen dramatically over the last few decades. In ILA studies, participants rate their experiences (eg, positive and negative affective states) repeatedly over time, often using electronic data collection via the internet or smartphones. The family of ILA methods encompasses real-time data collection as realized in ecological momentary assessments (EMAs) [1] and day-recall methods, including end-of-day (EOD) diaries [2] and the day reconstruction method (DRM) [3]. ILA methods offer several advantages compared with conventional types of assessment (eg, traditional questionnaires or clinical interviews). By inquiring about experiences that occurred over brief periods (eg, the past few minutes or the last day) in people's natural environments, ILA methods reduce recall bias and reliance on memory heuristics and can provide self-reports with improved ecological validity. In addition, the fine-grained data resulting from densely repeated assessments can be used to examine short-term, within-person processes that cannot be captured with traditional cross-sectional study designs [4,5].

ILA methods offer many novel insights into people's health and emotional functioning. For a long time, mental health research has predominantly focused on individuals' average levels of emotions. However, many aspects of people's emotional lives are not captured by how they feel on average. ILA adds a needed time dimension that allows assessment of the ebb and flow of subjective experiences and emotions. In many ways, ILA methods encourage a paradigm shift, changing the focus from emotions as static entities to studying them as dynamic processes [6]. For example, many mental health problems, including borderline personality disorder and bipolar disorder, are characterized by emotion dynamics, including increased intrapersonal *variability* or *instability* in affect [7-9]. Relatedly, the degree to which people's feelings are self-predictive or *linger* over time (described as emotional *inertia* or emotion *network density*) has been viewed as an important indicator of problems with successful emotion regulation [9-11] and as an important feature of mental health. Emotion dynamics may also be associated with physical health. The dynamic interplay of positive and negative affective states (the ability to experience positive and negative affective states in concert, *emotional dialecticism*, or *mixed emotions*) has been proposed as an indicator of emotional complexity and has been shown to benefit health outcomes [12,13].

An important advantage of ILAs is the ability to capture individual differences in emotion dynamics directly from repeated assessments. Researchers have enthusiastically embraced this possibility and have constructed measures of emotion dynamics from a range of ILA data sources, including EMAs, EOD diaries, and the DRM. EMAs have often been regarded as a *gold standard* among ILA methods in that respondents describe their momentary experiences as they are

happening in real time (or close to real time) [1]. However, EMAs are relatively expensive to implement and burdensome for participants. EOD diaries, wherein respondents complete a single rating at the end of each day (often using a 24-hour recall period), represent a less costly and more practical alternative that is frequently used for the study of emotion dynamics [9]. The DRM offers yet another alternative that affords a granular assessment of emotional experiences over the course of a day [3,14]. This method has respondents first revive memories of (ie, reinstantiate) the previous day by asking them to divide the day into episodes; for each episode of the day, respondents then provide information about what they were doing and rate their emotional experiences. Compared with EOD diaries and EMAs, the DRM has found less widespread attention in research on emotion dynamics [13]. However, the fact that respondents complete the DRM in a single session (eg, administered over the internet) makes it an attractive method for this purpose because it can be readily implemented with large samples to address population-level research questions [3].

It is often implicitly assumed that data from different ILA methods can be used interchangeably to construct measures of emotion dynamics, even though the specific emotion processes captured by the methods differ in possibly meaningful ways (Table 1). For instance, whereas EMA and the DRM collect information based on relatively brief time intervals between assessments (over periods of hours), EOD diaries provide 1 assessment per day. It is not at all clear that emotion regulation processes occur in similar ways across these different timescales [6]. In addition, whereas EMA protocols typically capture emotion fluctuations that occur both within and across days, EOD diaries target between-day variation, and the DRM is limited to within-day variation (given that the DRM is typically administered only for a single day). The reasons for day-to-day fluctuations in emotions may be conceptually quite different from those generating intraday fluctuations. Furthermore, compared with EMAs, both EOD diary and DRM ratings require retrospection, which can introduce recall bias. Memory heuristics in EOD diary recall (eg, reporting the most salient or peak experiences) have been shown to distort estimates of people's average experience levels to some extent [15,16], and the assessment of dynamic aspects of emotions may be similarly (or even more) distorted by these memory processes. Another consideration is that a lower number of data points per person negatively impacts the reliability of measures of emotion dynamics [17]; administering the DRM for a single day (ie, the previous day) puts an upper limit on the number of episodes obtained per person, whereas the number of assessments in EMAs and EOD diaries is determined by the research design. In addition, in contrast to EMAs and EOD diaries, the DRM leaves the selection of the number and temporal spacing of episodes to the respondents. This self-selection of episodes may introduce potential selection biases if the experiences respondents choose to report systematically differ from those that are not reported.

**Table 1.** Comparison of ecological momentary assessments, end-of-day diaries, and day reconstruction method measurement features.

Measurement characteristics	Intensive longitudinal assessment methods		
	EMA <sup>a</sup>	EOD <sup>b</sup> diary	DRM <sup>c</sup>
<b>Environmental or ecological context</b>			
Temporal dynamics covered	Within-day and across multiple days	Across multiple days	Within a single day
Spacing between targeted experiences	Several minutes to several hours	1 day	Several minutes to several hours
Timing of completion of assessments	Completed in succession over the course of each day	Completed in succession at the end of each day	All assessments completed in a single session
<b>Reliance on recall</b>			
Targeted reporting period	Moments or very short time periods	Overall summary of all experiences of the day	Summary of experiences during episodes of the day
Participant recall required	Moment (working memory)	Same day (episodic memory)	Previous day (episodic memory), with reinstatement
<b>Potential for selection bias</b>			
Selection of targeted experiences	Number and spacing of moments determined by prompting schedule	Fixed (daily)	Number and spacing of episodes determined by respondent
<b>Reliability of measures of emotion dynamics</b>			
Targeted number of measurement occasions	Dependent on prompting schedule (total of 42 in this study)	7 per week (total of 7 in this study)	Approximately 5 to 20 in total, determined by the respondent

<sup>a</sup>EMA: ecological momentary assessment.

<sup>b</sup>EOD: end-of-day.

<sup>c</sup>DRM: day reconstruction method.

Given that research on emotion dynamics derived from ILA is relatively new, researchers are using these measures without solid empirical knowledge about the impact that different ILA methods may have on their findings. To address this problem, this study aims to directly examine the extent to which common ILA methods produce comparable or dissimilar measures of intraindividual emotion dynamics. We used affect ratings collected using EMAs, EOD diaries, and DRM from the same individuals to derive many of the most commonly used measures of emotion dynamics from each method (including mean affect levels; affect variability, stability, and inertia; and emotion network density, emotional dialecticism, and mixed emotions) and examined the convergent validity of these measures across ILA methods. To address the question of whether the use of different ILA methods hampers the reproducibility of findings in health research, we further examined the extent to which measures of emotion dynamics constructed from each of the 3 methods demonstrate corresponding relationships with physical health outcomes (general health, pain, and fatigue). As the relationships between emotion dynamics and physical health are of particular interest in research on later points of the adult life span, this study focused on adults aged 50 years and older.

## Methods

### Participants and Procedures

The data for this study were collected as part of a larger project conducted in the Understanding America Study (UAS). The

UAS is a probability-based internet panel maintained at the University of Southern California (USC) Center for Economic and Social Research [18]. It comprises about 8000 panel members, including about 3500 respondents aged 50 years or older. In contrast to convenience (*opt-in*) panels, where people self-select to participate, UAS panel members are recruited through nation-wide, address-based sampling. UAS respondents without previous internet access are equipped with a tablet and broadband internet. In 2017, the UAS started building a data collection environment to enable ILA research with a nationally representative sample [19]. As is typical for large-scale internet panels, UAS panelists answer various surveys on a regular basis, but their level of participation typically does not rise to the magnitude of involvement required for ILAs. The goal of the original study was to demonstrate the feasibility of implementing ILA methods in older participants in this national internet panel. The data for this study were collected between July 2017 and September 2018 during the pilot waves of this project.

The study was approved by the USC Institutional Review Board. UAS panelists aged 50 years or older who were using iOS or Android mobile devices were eligible to participate in the pilot waves. Respondents were screened for current smartphone usage for the purposes of EMA data collection (about 30% were deemed ineligible because they did not use iOS or Android mobile devices). Eligible respondents were provided with information about the project and asked if they were willing to participate. Participants were selected randomly among eligible respondents who consented to participate, with a consent rate

of 86% (112/130). As part of the study, participants were asked to complete EMAs and EOD diary questions for 7 consecutive days and to complete the DRM for one of these days (randomly selected). The questions included 1 item addressing positive affect (PA; happiness) and 1 item addressing negative affect (NA; sadness or dejection) that were used for the analyses (additional emotion questions, such as cheerful, frustrated, angry, lonely, or relaxed, were only administered to a small subset of respondents in the DRM and are therefore not used here).

### EMAs

EMA data were collected on participants' own mobile phones with an app programmed using NubiS software. NubiS is an open-source, secure data collection, storage, and dissemination system developed by the Center of Economic and Social Research at USC. On each study day, respondents were prompted through cell phone beeps to complete EMA questions 6 times per day. Participants could specify the first and last possible prompt time for each day in the app before data collection, where the first prompt could be selected to occur between 6 AM and 11 AM and the last prompt between 7 PM and 11 PM of each day. Prompts were delivered using a stratified random sampling scheme that generated consecutive random prompts within the user-specified time window with a uniform probability between  $0.75 \times (\text{time window}/6)$  and  $1.10 \times (\text{time window}/6)$  hours (approximately between 1 and 3 hours) after the previous prompt. Each time the respondents received the prompt, they had 8 min to start answering questions. A reminder prompt was sent halfway during the 8-min time window if a participant did not immediately respond to an incoming prompt. Respondents were instructed not to respond to any prompt when they were driving and answer questions only if they were in a safe, secure, and private place. Each EMA survey included questions about location, social environment, and physical symptoms that were not analyzed here. The questions examined in this study were "before the prompt, how happy were you feeling" (PA), and "before the prompt, how dejected/blue/downhearted were you feeling" (NA). Responses were given on a 0 to 100 horizontal visual analog scale with anchors that ranged from *not at all* to *extreme*.

### EOD Diaries

Respondents were asked to complete a web-based EOD diary survey in the evening (after 6 PM) of each of the 7 study days. Respondents completed the daily diaries using their laptops or tablets in their UAS account. The EOD diary survey questions paralleled those used in the EMA reports. The PA question was, "please move the slider to represent how happy you felt today," and the NA question was, "please move the slider to represent how dejected/blue/downhearted you felt today." Responses were given on a 0 to 100 horizontal visual analog scale with anchors that ranged from *not at all* to *extremely*.

### DRM

The DRM was administered on the web once between days 2 and 7 of the study after participants had completed the daily diary questions. Participants were first asked when they woke up on the previous day and for how many hours they were

awake. Next, they were asked to "think of yesterday as a series of scenes in a movie" and to divide the day into episodes. Starting at the time they woke up, participants entered a label (using open-ended text entry) for the first episode that best described what they did during that time, specified an ending time for the episode, and clicked *next* to move to the next episode. This process was repeated for subsequent episodes until participants reached the end of the day. Participants were told that many people define episodes that last between 15 min and 2 hours, but they were encouraged to define as many episodes as made sense to them and could specify episodes of any duration. If participants were awake for more than 24 hours, they were asked to enter episodes for the first 24 hours they were awake. After the complete day had been reconstructed, participants completed questions about where they were, whom they interacted with, and how they felt during each episode. Consistent with the rating scale format used in the original DRM [3], emotions were rated on a 7-point scale from 0=*not at all* to 6=*very much*. Participants were asked to "please rate each feeling on the scale given," and the emotions presented included *happy* (PA) and *sad* (NA). As the number of scale points used in the DRM differed from the number of scale points used in EMAs and EOD diaries, the DRM scores were transformed into a 0 to 100 scale using the following formula: transformed score =  $100 \times (\text{original score} + 0.5) / 7$  [20].

### Construction of Measures of Emotion Dynamics

The selection of measures of emotion dynamics was guided by previous reviews of the most commonly employed measures in applied research settings [9,11,21]. All measures were calculated in the same way for each ILA method based on each respondent's PA and NA ratings across moments (EMAs), days (EOD diaries), or episodes (DRM). To be included in the analyses, we required that a participant provide at least 4 observations for each of the ILA methods, which was deemed the minimum number of observations required to reasonably compute the various measures.

#### Mean PA and NA Levels

Measures of an individual's mean PA and NA levels represent the most well-known and prominent indicators of psychological well-being. For each ILA method, they were calculated by taking the average of each respondent's ratings across assessment time points, separately for PA and NA.

#### Variability of PA and NA

Emotion variability captures the magnitude (or amplitude) of fluctuations of a person's emotional states around the person's average emotion level [22]. Measures of variability were created by calculating the within-person SD of each respondent's PA and NA ratings across assessment time points.

#### Instability of PA and NA

Emotional instability measures differ from variability measures in that they explicitly consider the temporal ordering of affective states. Specifically, emotional instability refers to the magnitude of shifts in PA and NA levels across consecutive assessments [7]. The root mean square successive difference measure was used for this purpose, which was calculated by taking the

average of the squared differences between successive ratings and taking the square root of this average for each respondent.

### ***Inertia of PA and NA***

The concept of emotional inertia refers to the degree to which an individual's emotions are resistant to change, such that levels of PA or NA persistently carry on over time. Inertia is commonly operationalized using the first-order autocorrelation of consecutive measurements in a time series to capture the temporal dependency of a person's PA and NA ratings [10]. Correspondingly, we obtained measures of inertia in PA and NA as the person-specific autoregressive slope in regression models in which an affect rating at one time point predicts the rating of the same affect item at the subsequent measurement time point.

### ***Emotion Network Density***

The concept of emotion network density is an extension of the inertia concept to multiple affect items. It evaluates the degree to which multiple emotions predict each other over time, reflecting the extent to which a person's overall system of PA and NA states is resistant to change. Whereas measures of inertia are calculated separately for each affect item, network density combines the temporal dependencies of multiple affect items in a single measure [23,24]. For 2 emotion items, it comprises the sum of the absolute value of 4 lagged parameters in a vector autoregressive model involving 2 autoregressive and 2 cross-lagged parameters of PA and NA. Specifically, we obtained the person-specific (autoregressive and cross-lagged) parameters of PA and NA (entered simultaneously as within-person centered predictors) in regression models in which either PA or NA served as the outcome variable. The network density measure was then created by taking the absolute value of the 4 regression parameters and calculating the sum of these absolute values separately for each person.

### ***Mixed Emotions***

The experience of mixed (or bittersweet) emotions has been conceptualized as the extent to which both PA and NA are felt together at the same point in time. The construct of mixed emotions involves the simultaneous activation (or co-occurrence) of both positive and negative experiences [13]. Simply knowing whether an individual has high average levels of PA and NA over time may tell us nothing about whether they experienced a blending of PA and NA at any given point in time [25]. Following previous research, a measure of mixed emotions was calculated using the MIN index [13,25,26]. The index is based on the ambivalence metric proposed by Kaplan [27], which is defined as *total affect* (the sum of ratings for PA and NA) minus *polarity* (the absolute difference between ratings for PA and NA) for a given time point. Arithmetically, this formula is equivalent to taking (2 times) the smaller value of the PA and NA ratings (ie, MIN [PA, NA]) at a given time point [28], such that the MIN index is high only if both emotions co-occur at high levels at that time point. The MIN index was calculated for each time point and averaged across the assessment time points of each respondent.

### ***Emotional Dialecticism***

The concept of emotional dialecticism (the reverse of the concept of affective bipolarity) refers to the degree to which individuals tend to experience PA and NA independently from each other rather than as bipolar opposites [29,30]. Although conceptually similar to mixed emotions, the measure is calculated as the within-subject correlation between PA and NA ratings of each respondent [29]. As the correlation between PA and NA is expected to be negative on average, more strongly negative values (ie, values approaching a correlation of  $-1.0$ ) on the measure represent less dialecticism, and less strongly negative values (ie, values approaching or exceeding a correlation of 0) represent more dialecticism.

Several additional considerations regarding the calculation of the measures of emotion dynamics are noteworthy. First, measures that take the temporal ordering into account (ie, measures of emotional instability, inertia, and emotion network density) require that the time intervals between 2 consecutive measurements are approximately equal. For this reason, when calculating these measures from EMA data, time periods from the evening of one day to the morning of the next day were omitted, as were momentary ratings that occurred after a time gap of more than 10 hours. Similarly, consecutive DRM ratings that were more than 10 hours apart from each other (measured from the midpoint of one episode to the midpoint of the next episode) were omitted when calculating these measures. Second, measures of emotion dynamics that involve lagged within-person relationships (ie, measures of inertia and emotion network density) are prone to high imprecision because of sampling error unless the number of measurement occasions is very large [17]. As the number of occasions in this study was relatively modest (especially for daily diaries), we used multilevel models to obtain empirical Bayes estimates of the person-specific regression parameters for calculating the inertia and network density measures. Compared with parameters from regression models that are estimated separately per respondent, the multilevel approach yields more precise parameter estimates for individual respondents [31].

### ***Health Variables***

Self-reported general health, pain, and fatigue were measured at the end of each of the 7 study days as part of the EOD diary assessments. For each health variable, the 7 scores were averaged into a summary measure for each person. Self-reported health was measured with the question "How was your health today?" with response options from the Short Form 36 general health item (excellent, very good, good, fair, or poor) [32]. To assess pain severity and fatigue levels, participants were asked to rate the following statements, *your average level of bodily pain* (no pain at all to extreme pain) and *how fatigued* (ie, *weary or tired*) *you were* (no fatigue at all to extreme fatigue), using a 0 to 100 horizontal visual analog scale.

### ***Statistical Analysis***

Data analyses were performed in 3 broad steps. The first step of the analysis examined the extent to which the 3 ILA methods (ie, EMAs, EOD diaries, and DRM) yielded comparable scores for each emotion dynamic. To accomplish this, we compared

the means and examined the correlations of measures addressing the same emotion dynamic across the ILA methods. Differences in mean scores were tested using analysis of variance (ANOVA) procedures with an ILA method as a within-person (ie, repeated measures) factor. An omnibus test of overall mean differences between the 3 methods was conducted first, followed by Bonferroni-adjusted pairwise tests of mean differences between the methods. For correlations of emotion dynamic measures between ILA methods, we roughly considered correlations of 0 to 0.35 as low, 0.36 to 0.67 as moderate, 0.68 to 0.89 as high, and 0.90 to 1.00 as indicating very high correspondence, following conventions [33]. To test whether the correlations differed between pairs of ILA methods (such that some ILA methods more highly correspond with each other than with other methods), we used Fisher z-transformed correlation coefficients and conducted Wald tests for differences in dependent correlations as implemented in Mplus version 8.4 (Muthén & Muthén) [34].

Whereas the first analysis step considered each measure of emotion dynamics separately, the second set of analyses examined how the measures are interconnected within each ILA method. Different measures of emotion dynamics may positively or negatively relate with each other in complex ways, and these relationships can be viewed to form a system or network of interdependencies among different features of emotional experience. Thus, evidence that the pattern of associations among the various measures corresponds across ILA methods would support the convergent validity of the methods. To examine this, we first inspected the structure of interconnections graphically using network visualization in the R package qgraph [35]. This technique represents a correlation matrix as a network in which each measure of emotion dynamics is represented as a node, and their interconnections are shown as edges between the nodes, allowing for a visual comparison of the correlation networks between ILA methods. To quantify the similarity of the correlation networks across ILA methods, we capitalized on centrality indices used in network analysis. In centrality analysis, the emotion dynamics are ordered in terms of the degree to which they occupy a central place in the overall network and exhibit many strong associations. We focused on the *simplest centrality metric, node strength*, which was calculated for each emotion dynamic as the sum of its absolute correlations with all other emotion dynamics in the network [36]. The ordering of the emotion dynamics' centralities was then descriptively compared across the ILA methods.

The third step examined the extent to which measures of emotion dynamics constructed from the different ILA methods demonstrate corresponding correlations with the health outcomes (general health, pain, and fatigue). To statistically compare the correlations of each emotion dynamic across the different ILA methods, we conducted Wald tests for differences in dependent Fisher  $r$ - to z-transformed correlation coefficients. In addition, effect sizes (Cohen  $q$ , where values of 0.1, 0.3, and 0.5 can be interpreted as small, medium, and large effects, respectively) were calculated to quantify the magnitude of differences in the correlations between ILA methods.  $P$  values  $<.05$  were considered significant for all analyses.

## Results

### Descriptive Characteristics

A total of 100 participants completed assessments using all 3 ILA methods, with 10 participants being excluded from the analyses because they provided less than 4 observations for at least one of the ILA methods (1 participant had fewer than 4 observations for all ILA methods, 3 for EMA, and 6 for the DRM), resulting in an analysis sample of 90 participants. The mean age of the analyzed sample was 62.4 years ( $SD$  7.7; range 51-87 years), 57% (51/90) were female, and 63% (57/90) were married. The sample was predominantly White (72/90, 80%) and non-Hispanic (83/90, 92%). The median household income was in the category between US \$60,000 and US \$74,999, and 57% (51/90) had a college degree. About half (42/90, 47%) were currently working, and about one-third (32/90, 36%) indicated that they were retired. Participants who were excluded from the analyses were less likely to hold a college degree ( $P=.04$ ) but otherwise did not differ from the analysis sample on these demographics.

In terms of participants' compliance with the ILA protocol, the mean number of completed EMA prompts per person in the analyzed sample was 29.9 ( $SD$  11.2; median 34) out of 42 possible ratings (6 per day across 7 days), yielding an average EMA completion rate of 71.2% (29.9/42). EOD diaries were, on average, completed on 6.9 ( $SD$  0.5; median 7) out of the 7 days, yielding an average completion rate of 99.0% (6.9/7). For the DRM, the mean number of episodes provided per person was 11.8 ( $SD$  5.7; median 11.0), which is roughly comparable with previous research using the DRM [3].

As the expected values of emotion dynamics that involve lagged within-person associations depend on the length of the lag (ie, autocorrelations are expected to be, on average, lower, the longer the lag time is), descriptive statistics for the time distances between ratings for each ILA method were also examined. For EMA prompts, the average distance between assessments (after elimination of overnight gaps and gaps  $>10$  hours) was 2.13 hours ( $SD$  0.89; median 1.94), with an IQR of 1.70 to 2.18 hours. The average time distance between consecutive EOD diaries was 24.01 hours ( $SD$  2.16; median 24.05; IQR 23.32-24.76 hours). For the DRM, the average distance between the midpoints of consecutive episodes was 1.68 hours ( $SD$  1.25; median 1.25; IQR 0.58-2.25 hours).

To evaluate the proportion of the total variance in affect that was accounted for by stable, trait-like differences as opposed to within-person fluctuations, intraclass correlations of PA and NA ratings were calculated for each ILA method, computed as the ratio of between-person variance to total (sum of within- and between-person) variance in ratings. The intraclass correlations were 0.63 (PA) and 0.54 (NA) for EMA reports; 0.57 (PA) and 0.53 (NA) for EOD reports; and 0.60 (PA) and 0.51 (NA) for DRM reports. Thus, between 37% (151.4/403.8) and 43% (175.6/410.9) of the variance in PA was within-person, and between 46% (142.6/311.2) and 49% (121.2/249.2) of the variance in NA was within-person, with consistency across the 3 ILA methods.

## Correspondence of Measures of Emotion Dynamics Across the ILA Methods

ANOVA models yielded no significant differences between the ILA methods for measures capturing mean levels in PA and NA, PA variability, and PA instability, whereas significant method differences were evident for the remaining measures of emotion dynamics (Table 2). In pairwise comparisons between the ILA methods, the averages of measures from EMAs and EOD diaries did not significantly differ from each other, with one exception: the average inertia for NA was smaller for EOD diaries than for EMAs. This difference is consistent with expectations, given a longer lag time between EOD diary assessments compared with the lag times between EMAs. For DRM-derived measures, the averages were less comparable with those from the other ILA methods. The DRM yielded higher averages of inertia for PA and NA, and a higher emotion network density, compared with both EMAs and EOD diaries (consistent with the shorter time lags of the DRM). In addition, the DRM yielded lower average scores for NA variability and NA instability and higher scores for mixed emotions and emotional dialecticism compared with both EMAs and EOD diaries (Table 2).

The correlations between measures derived from EMAs and EOD diaries suggested very high correspondence ( $\rho > 0.90$ ) between these 2 methods for individuals' mean PA levels, mean NA levels, and mixed emotions, and high correspondence ( $\rho = 0.68$  and  $0.80$ ) for measures of PA and NA variability (Table 3). EMA- and EOD diary-derived measures further showed moderate-to-high correspondence ( $\rho$  ranging between 0.51 and 0.70) for PA and NA instability and emotional dialecticism measures, and low-to-moderate correspondence ( $\rho$  ranging between 0.17 and 0.57) for measures of PA and NA inertia and emotion network density. For most of the measures derived from the DRM, the correlations with both EMAs and EOD diaries were significantly weaker by comparison (Table 3). Specifically, the DRM showed high correspondence with EMAs and EOD diaries ( $\rho$  ranging between 0.74 and 0.80) for individuals' mean PA levels, mean NA levels, and mixed emotions, moderate correspondence ( $\rho$  between 0.38 and 0.54) for PA and NA variability measures, low-to-moderate correspondence ( $\rho$  between 0.19 and 0.41) for PA and NA instability measures, and low correspondence ( $\rho$  between 0.03 and 0.32) for the remaining measures (PA and NA inertia, emotion network density, and emotional dialecticism).

**Table 2.** Mean (SD) of measures of emotion dynamics for each assessment method.

Emotion dynamic	Assessment method, mean (SD)			ANOVA <sup>a</sup> of mean differences	
	EMA <sup>b</sup>	EOD <sup>c</sup> diary	DRM <sup>d</sup>	<i>F</i> value ( <i>df</i> =2178)	<i>P</i> value
PA <sup>e</sup> mean level	66.24 (16.1)	68.40 (16.1)	66.79 (17.6)	2.61	.08
NA <sup>f</sup> mean level	12.60 (13.3)	12.12 (14.7)	14.00 (11.6)	2.73	.07
PA variability	11.29 (4.6)	11.65 (6.7)	11.93 (7.8)	0.42	.66
NA variability	9.05 <sup>g</sup> (7.3)	9.02 <sup>g</sup> (9.3)	5.79 <sup>h</sup> (8.5)	10.14	<.001
PA instability	12.92 (6.3)	14.51 (9.1)	13.53 (9.6)	1.16	.32
NA instability	10.41 <sup>g</sup> (8.8)	11.08 <sup>g</sup> (11.6)	6.58 <sup>h</sup> (9.8)	10.13	<.001
PA inertia	0.41 <sup>g</sup> (0.1)	0.38 <sup>g</sup> (0.1)	0.46 <sup>h</sup> (0.2)	8.72	<.001
NA inertia	0.34 <sup>g</sup> (0.1)	0.26 <sup>h</sup> (0.1)	0.42 <sup>i</sup> (0.2)	32.49	<.001
Network density	0.86 <sup>g</sup> (0.2)	0.91 <sup>g</sup> (0.2)	0.99 <sup>h</sup> (0.3)	10.23	<.001
Mixed emotions	10.90 (10.6)	10.03 <sup>g</sup> (11.6)	12.42 <sup>h</sup> (9.5)	6.24	.002
Dialecticism	-0.34 <sup>g</sup> (0.3)	-0.36 <sup>g</sup> (0.2)	-0.21 <sup>h</sup> (0.2)	13.14	<.001

<sup>a</sup>ANOVA: analysis of variance.

<sup>b</sup>EMA: ecological momentary assessment.

<sup>c</sup>EOD: end-of-day.

<sup>d</sup>DRM: day reconstruction method.

<sup>e</sup>PA: positive affect.

<sup>f</sup>NA: negative affect.

<sup>g,h,i</sup>Means in the same row with different superscripts differ significantly from each other using Bonferroni-adjusted pairwise comparisons and means in the same row without superscripts or that share the same superscript do not differ.

**Table 3.** Correlations of emotion dynamics across assessment methods.

Emotion dynamic	Correlation coefficients			Difference between correlations	
	EMA <sup>a</sup> -EOD <sup>b</sup> diary	EMA-DRM <sup>c</sup>	EOD diary-DRM	$\chi^2$ ( <i>df</i> =2)	<i>P</i> value
PA <sup>d</sup> mean level	0.95 <sup>e</sup>	0.80 <sup>f</sup>	0.79 <sup>f</sup>	50.52	<.001
NA <sup>g</sup> mean level	0.95 <sup>e</sup>	0.78 <sup>f</sup>	0.74 <sup>f</sup>	70.68	<.001
PA variability	0.68 <sup>e</sup>	0.54	0.38 <sup>f</sup>	15.46	<.001
NA variability	0.80 <sup>e</sup>	0.47 <sup>f</sup>	0.45 <sup>f</sup>	25.68	<.001
PA instability	0.51 <sup>e</sup>	0.33	.019 <sup>f</sup>	9.05	.01
NA instability	0.70 <sup>e</sup>	0.37 <sup>f</sup>	0.41 <sup>f</sup>	14.99	<.001
PA inertia	0.46 <sup>e</sup>	0.18	.017 <sup>f</sup>	6.43	.04
NA inertia	0.17	0.26	0.03	3.33	.19
Network density	0.36	0.32	0.26	0.77	.68
Mixed emotions	0.92 <sup>e</sup>	0.76 <sup>f</sup>	0.76 <sup>f</sup>	30.35	<.001
Dialecticism	0.57 <sup>e</sup>	0.17 <sup>f</sup>	0.31	14.25	<.001

<sup>a</sup>EMA: ecological momentary assessment.

<sup>b</sup>EOD: end-of-day.

<sup>c</sup>DRM: day reconstruction method.

<sup>d</sup>PA: positive affect.

<sup>e,f</sup>Correlations in the same row with different superscripts differ significantly from each other using Bonferroni-adjusted pairwise comparisons and means in the same row without superscripts or that share the same superscript do not differ. Correlations >0.21 are significant at  $P<.05$ , correlations >0.28 are significant at  $P<.01$ , and correlations >0.35 are significant at  $P<.001$ .

<sup>g</sup>NA: negative affect.

The goal of the primary analyses was to examine the comparability of measures that were computed in the way they would most likely be computed in applied research, that is, based on all available data for each ILA method. However, given that the DRM captures only 1 single day, an interesting question is whether the correspondence between EMA-based and DRM-based measures increases when measures of emotion dynamics are derived from EMAs for the exact day before the DRM was completed (ie, the exact day respondents were asked to rate in the DRM). Secondary analyses were conducted to address this question (based on 71 participants who had at least 3 EMA reports for the day before the DRM). The pattern of correlations with DRM-based measures was highly similar when measures of emotion dynamics were derived from all EMA data versus same-day EMA data (the median Cohen  $q$  for the difference in correlations was  $-0.014$ ).

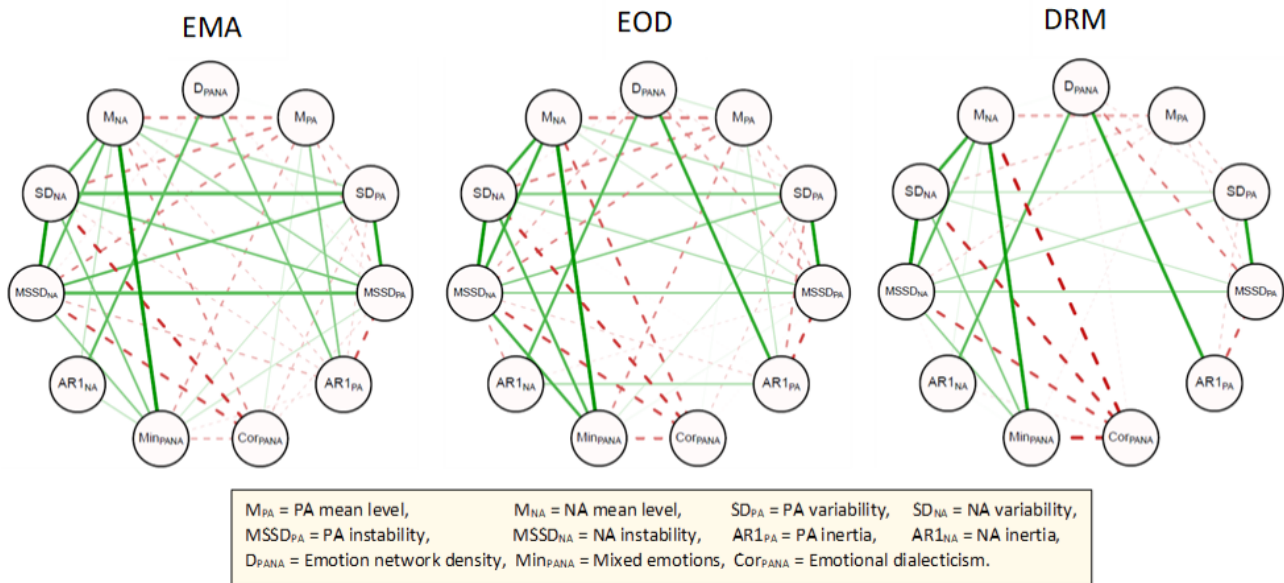
### Patterns of Associations Among Emotion Dynamics Within Each ILA Method

The network of pairwise associations among the various emotion dynamics is illustrated in [Figure 1](#) for each ILA method.

Examining the network of EMA-derived measures, several patterns are noteworthy. First, moderate-to-strong positive connections were evident among all variability and instability measures ( $\rho\geq 0.67$ , indicated by thick green lines). Second, the measure of mean NA levels showed positive connections with NA variability and NA instability and mixed emotions ( $\rho\geq 0.67$ ). Third, although measures of mean PA levels and emotional dialecticism were weakly correlated with each other ( $\rho=0.33$ ), both showed moderate-to-strong negative connections with mean NA levels, NA variability, and NA instability ( $\rho\leq -0.54$ , indicated by thick red lines). Finally, emotion network density was positively associated with PA inertia and NA inertia ( $\rho\geq 0.59$ ), whereas all 3 measures showed otherwise few connections in the network. Visual comparison of the correlation networks suggests that the patterns of association were similar for measures derived from EMAs and EOD diaries. However, DRM-derived measures generally showed fewer and weaker interconnections ([Figure 1](#)).



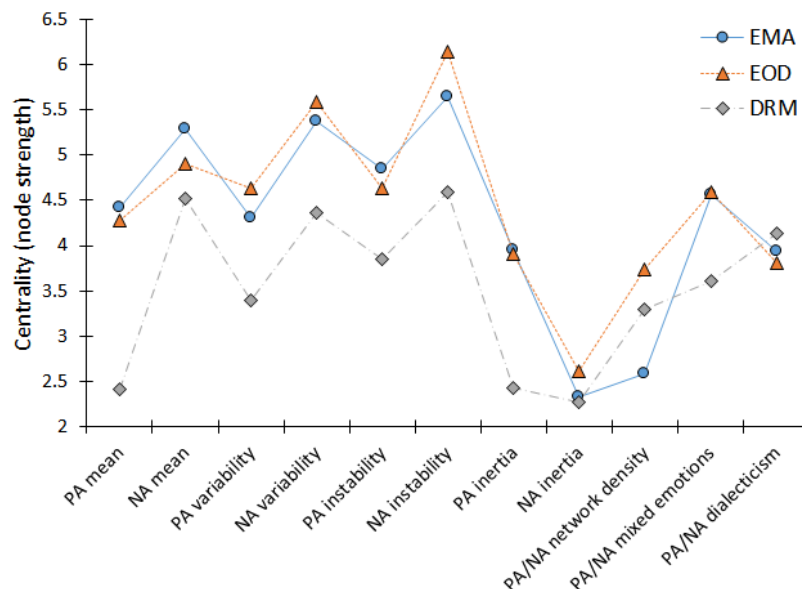
**Figure 1.** The network of pairwise correlations among emotion dynamics for each assessment method. Green lines represent positive relations, and red dashed lines represent negative relations. The thickness and transparency of the lines correspond with the magnitude of correlations. For clarity, lines are shown only for correlations that surpass the  $P < .05$  significance threshold (above an absolute value of 0.21). EMA: ecological momentary assessment; EOD: end-of-day; DRM: day reconstruction method; PA: positive affect; NA: negative affect.



The magnitude of the interconnections is quantified in the node strength centralities of the measures (Figure 2). For emotion measures derived from EMA, those tapping emotion levels, variability, and instability occupied the most central places with the strongest interconnections; measures for NA showed consistently greater centralities than corresponding measures for PA. As can be seen in Figure 2, the ordering of node strengths corresponded very closely between EMA measures

and those derived from EOD diaries, with a correlation of  $\rho = 0.92$  between these 2 ILA methods. The node strength centralities of DRM-derived measures were generally lower (corresponding with weaker interconnections among DRM-derived measures), and the ordering of node strengths showed moderate-to-high correspondence with those from EMAs ( $\rho = 0.75$ ) and EOD diaries ( $\rho = 0.70$ ).

**Figure 2.** Centrality (node strength) of emotion dynamics for each assessment method. EMA: ecological momentary assessment; EOD: end-of-day; DRM: day reconstruction method; PA: positive affect; NA: negative affect.



**Relationships With Health Outcomes**

Finally, relationships between the measures of emotion dynamics and health outcome variables were examined. Overall, the emotion dynamics showed small- to medium-sized correlations with self-reports of general health and fatigue, and

somewhat less pronounced correlations with pain (Figure 3). Comparisons of the correlation coefficients between ILA methods yielded very few differences: out of 33 Wald tests (for 11 emotion dynamics and 3 health outcomes), only 1 indicated a significant difference between the ILA methods (Wald

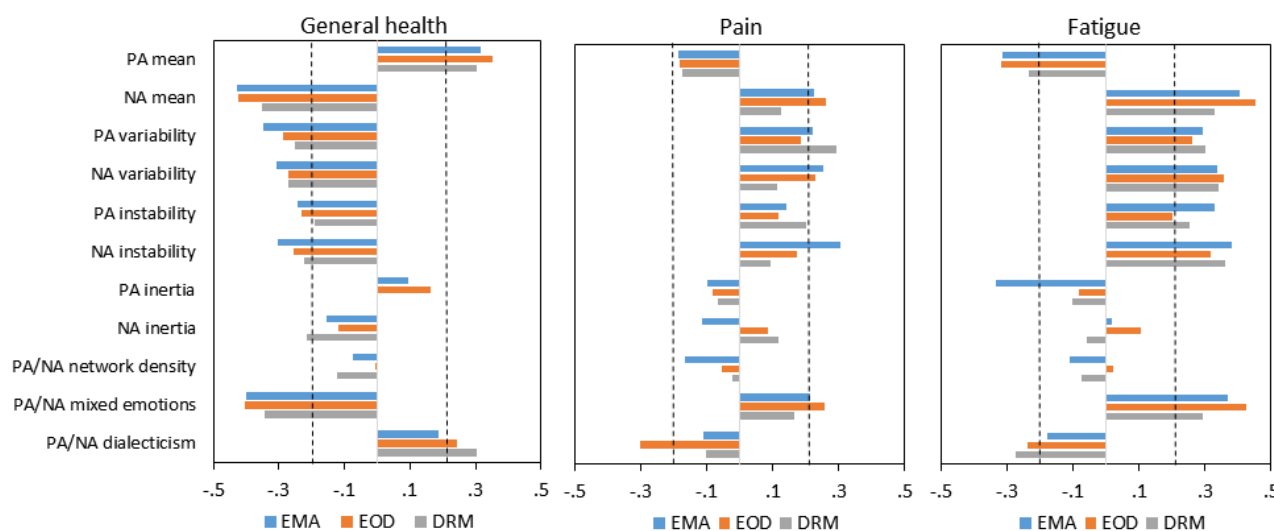
$\chi^2=6.19$ ,  $P=.04$  for the correlations between PA inertia and fatigue; post hoc pairwise comparisons showed that the correlation was significantly more negative for EMA than for EOD diary reports,  $z=-2.38$ ,  $P=.02$ ). Average effect sizes (absolute values of Cohen  $q$ ) for the differences in correlations between the ILA methods were very small for EMAs versus EOD diaries (general health, mean absolute  $q=0.04$ ; pain, mean absolute  $q=0.08$ ; and fatigue, mean absolute  $q=0.09$ ), EMAs versus DRM (general health, mean absolute  $q=0.07$ ; pain, mean absolute  $q=0.10$ ; and fatigue mean absolute  $q=0.08$ ), and for EOD diaries versus DRM (general health, mean absolute  $q=0.07$ ; pain, mean absolute  $q=0.09$ ; and fatigue, mean absolute  $q=0.08$ ).

As shown in Figure 3, for all ILA methods, the most pronounced relationships with health outcomes were evident for mean emotion levels, variability, and instability measures in expected directions. Higher mean PA levels were consistently correlated with better general health, less pain, and less fatigue, whereas higher mean NA, variability (PA and NA), and instability (PA and NA) correlated with poorer general health, more pain, and more fatigue. Higher values of emotional dialecticism were associated with better health outcomes, whereas, contrary to

theoretical expectations [37], more mixed emotions were associated with worse health outcomes. Finally, inertia (PA and NA) and emotion network density measures showed almost no relationship with the health outcomes.

We also explored whether measures of emotion dynamics demonstrated incremental validity in predicting the health outcomes above and beyond average emotion levels. To examine this, a series of multiple regression analyses were conducted in which the health outcomes were regressed on a measure of emotion dynamics after controlling for PA and NA mean levels separately for each ILA method. In these models, the only significant predictor of general health was the mean level of NA, consistently for each ILA method ( $P<.05$  in all instances). For pain as an outcome variable, no emotion measure (including mean PA and NA) was a significant predictor in the multiple regressions; the exception was that PA variability uniquely predicted pain in the DRM ( $P=.02$ ). For fatigue, the only significant predictor was the mean level of NA, again consistently for each ILA method ( $P<.05$  in all instances). In addition, PA variability uniquely predicted fatigue in the DRM ( $P=.03$ ).

**Figure 3.** Correlations of measures of emotion dynamics with self-reports of general health, pain, and fatigue for each assessment method. Horizontal bars represent the magnitude and direction of correlation coefficients. Correlations exceeding a value of 0.21 (indicated by vertical dashed lines) are significant at  $P<.05$ . EMA: ecological momentary assessment; EOD: end-of-day; DRM: day reconstruction method; PA: positive affect; NA: negative affect.



## Discussion

Emotions fluctuate and change over time, and ILA methods are uniquely suited to quantify the temporal dynamics of people’s emotional lives. There has been a surge of interest in creating measures that tap a variety of emotion dynamics from ILA. To our knowledge, this is the first study to directly compare EMA, EOD diary, and DRM ILA methods in the measurement of intraindividual emotion dynamics. If different ILA methods produce noncorresponding measures of emotion dynamics, this would have important research implications in that it would question the validity of the measures and threaten the reproducibility of empirical research results.

EMAs and EOD diaries are arguably the 2 most commonly used ILA methods [4]. Previous research has documented that

individual differences in mean levels of positive and negative experiences are highly correlated between these methods [38,39]. This finding was confirmed in this study. Expanding on previous research, we found that several measures of emotion dynamics derived from EOD diaries also reproduced those derived from EMAs very well. Specifically, measures of emotion variability, instability, mixed emotions, and emotional dialecticism showed substantial correspondence between the 2 methods, with comparable mean levels and moderate-to-high correlations. Considering the argument that momentary and daily fluctuations in emotions may be quite different conceptually, the level of agreement between the methods is perhaps somewhat surprising. For example, given that EMAs encompass the sum of within-day and between-day sources of intraindividual variation, whereas EOD diary ratings are limited to between-day variation, one might expect the average

variability to be higher in EMAs than in EOD diaries, which we did not find. Our results suggest that differences in the time scale and frequency of measurement inherent in EMA and EOD diary ratings do not dramatically impact measures of emotion variability, instability, mixed emotions, and emotional dialecticism.

However, this does not mean that EMA and EOD diary measures can be viewed as universally interchangeable. Measures of inertia and emotion network density were correlated at levels below 0.50 between the methods. In contrast to the other measures of emotion dynamics, inertia and emotion network density measures are specifically focused on temporal dependencies between successive self-report ratings; that is, they capture the rate (or speed) of changes rather than the magnitude of changes. This suggests that the measurement of how emotions evolve over time and at what rate may not be captured in the same way by EOD diaries and EMAs.

The DRM was originally developed as an alternative to EMAs for use in large and population-representative samples, yet few previous studies have directly compared the information obtained from the DRM with that obtained from EMAs [40-42] or EOD diaries [43,44]. In this study, the DRM demonstrated high correspondence with other ILA methods for measures of average emotion levels and mixed emotions and moderate correspondence for emotion variability. For more complex emotion dynamics, the DRM showed low correlations with EMAs and EOD diaries. Similarly, the networks of pairwise associations among the various emotion dynamics were only moderately concordant between the DRM and the other ILA methods. This suggests that the DRM may adequately capture people's average emotion levels (and, to some extent, their emotion variability) but may not serve as a direct replacement of other ILA methods in research on more intricate emotion dynamics.

Interestingly, all ILA methods produced very similar patterns of relationships with health outcomes. Higher mean PA and lower mean NA levels were associated with better physical health outcomes with medium (general health and fatigue) and small-to-medium (pain) effect sizes, replicating previous research on these mind-body relationships [45-47]. The consistency of these results across ILA methods suggests that each of them can equally contribute to understanding the linkages between people's general emotion levels and health outcomes in older adulthood.

Measures of emotion variability and more emotional instability derived from each ILA method were also consistently related to health outcomes in expected directions. These results are noteworthy considering that even though the last decades have witnessed a surge in research linking emotion variability with maladaptive outcomes, most of this research has focused on psychological well-being (eg, depression, anxiety) rather than physical health outcomes. In a large meta-analysis, Houben et al [9] found small-to-medium effect sizes for relationships between psychological well-being and measures of emotion variability ( $\rho=0.18$ ) and instability ( $\rho=0.21$ ). Our results suggest effect sizes of similar magnitude for physical health outcomes.

Measures of mixed emotions (MIN index) and emotional dialecticism (within-person correlation of PA and NA) were moderately negatively correlated with each other and showed opposite relationships with the health outcomes within each of the ILA methods. This finding would appear counterintuitive, given that the 2 measures aim at capturing conceptually similar concepts. However, previous studies have found a generally weak correspondence between these indices of mixed emotions and emotional dialecticism [13,26], suggesting that they measure different aspects of emotional experience. Consistent with our results, higher emotional dialecticism scores have previously been shown to predict fewer health symptoms [12], whereas mixed emotions assessed with the MIN index were associated with more physical disability across adulthood [13].

Measures that focus on temporal dependencies of emotional states (emotional inertia and emotion network density) were practically uncorrelated with the physical health outcomes in this study, contrary to previous meta-analytic findings that higher emotional inertia relates to worse psychological well-being, albeit with overall small effect sizes [9]. It is possible that temporal dependency measures play a lesser role in understanding physical health compared with mental health. However, previous simulation studies have also demonstrated that measures of inertia derived from ILAs tend to have very low reliability, and this may have attenuated the observed correlations with health outcomes. A recent Monte Carlo simulation by Du and Wang [17] suggests that even if the emotional states themselves are assessed with near-perfect reliability, up to 100 measurement occasions per person may be necessary to obtain reliabilities  $>0.70$  for inertia measures, whereas individual differences in mean levels, variability, and instability measurement required substantially fewer measurement occasions to be reliably measured in the simulation study. Future research would benefit from implementing ILA methods across multiple waves (eg, using measurement bursts) as a means to estimate the test-retest reliability of measures of emotion dynamics and to correct their correlations with health outcomes for unreliability in empirical samples.

It is also noteworthy that once the explanatory power of mean levels of PA and NA was taken into account, measures of emotion dynamics showed little to no added value in the prediction of the health outcomes, regardless of the ILA method. This finding corresponds with recent findings by Dejonckheere et al [21], suggesting that more complex emotion dynamics add little to the prediction of psychological well-being and emotion disorders after controlling for mean affect levels. Our findings suggest that caution is similarly warranted when assuming that measures of complex emotion dynamics have unique predictive utility for understanding physical health parameters, even though this would need to be confirmed in larger studies.

Several limitations of this study need to be considered. First, the study was restricted to older individuals aged 50 years and above, and the results may not be generalizable to younger adults. However, evidence for convergent validity across ILA methods in this age group may be particularly important given that older people may have more problems with electronic ILA data collection and that potential cognitive problems in this age group may interfere with accurate self-reports [48]. Second, the

sample size of 90 respondents was relatively modest, although it was comparable with previous studies examining the correspondence of EMAs with EOD diaries [38] or the DRM [40,41]. A third limitation is that whereas EMAs and EOD diaries used the same wording of affect items (ie, happy, dejected/blue/downhearted) and the same response scale (a 0-100 visual analog scale), the DRM used a partially different wording (ie, happy, sad) and different response scale (a 7-point numeric response scale) to collect affect ratings. This may have artificially deflated the correspondence between measures of emotion dynamics derived from the DRM and the other ILA methods. Holding the items and response scales constant across ILA methods in future studies would enhance the rigor of comparisons by minimizing the potential impact of such extraneous factors. Fourth, it is also possible that completing multiple EMA ratings throughout the day impacted participants' EOD diary and DRM affect ratings, which may have artificially inflated the correspondence between the different measures. Previous literature found little evidence that recall ratings are impacted by momentary reporting [49], but we did not specifically examine this possibility here. Finally, even though we examined a variety of measures of emotion dynamics, the

study was limited to measures that can be derived from 2 affect questions. Additional measures of emotion dynamics that have been proposed in the literature were not considered here because they require administration of multiple PA or NA items at each measurement occasion (examples are emotional granularity, which captures the extent to which individuals differentiate between multiple emotions of the same valence [50], and emodiversity, which captures the extent to which individuals experience a narrow or wide range of different emotions [51]).

In summary, EMAs and EOD diaries correspond moderately to highly with each other in the information they provide about individual differences in various emotion dynamics. Compared with these ILA methods, the DRM provides corresponding information about emotion levels and (to a lesser extent) variability, but not about more complex emotion dynamics. Our results caution researchers against viewing these ILA methods as universally interchangeable. Although measures of emotion dynamics derived from all ILA methods showed small-to-moderate relationships with physical health outcomes, the unique predictive ability of more complex emotion dynamics for understanding health outcomes remains to be established.

## Acknowledgments

This study was supported by grants from the National Institute on Aging (R21 AG061364 and U01 AG054580) and the Social Security Administration. The sponsors were not involved in the conduct of this study or the preparation of this paper.

## Conflicts of Interest

AS is a senior scientist with the Gallup Organization and a consultant with Adelphi Values. All other authors declare no conflicts of interest.

## References

1. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol* 2008;4:1-32. [doi: [10.1146/annurev.clinpsy.3.022806.091415](https://doi.org/10.1146/annurev.clinpsy.3.022806.091415)] [Medline: [18509902](https://pubmed.ncbi.nlm.nih.gov/18509902/)]
2. Bolger N, Davis A, Rafaeli E. Diary methods: capturing life as it is lived. *Annu Rev Psychol* 2003;54:579-616. [doi: [10.1146/annurev.psych.54.101601.145030](https://doi.org/10.1146/annurev.psych.54.101601.145030)] [Medline: [12499517](https://pubmed.ncbi.nlm.nih.gov/12499517/)]
3. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: the day reconstruction method. *Science* 2004 Dec 2;306(5702):1776-1780. [doi: [10.1126/science.1103572](https://doi.org/10.1126/science.1103572)] [Medline: [15576620](https://pubmed.ncbi.nlm.nih.gov/15576620/)]
4. Schneider S, Stone AA. Ambulatory and diary methods can facilitate the measurement of patient-reported outcomes. *Qual Life Res* 2016 Mar;25(3):497-506 [FREE Full text] [doi: [10.1007/s11136-015-1054-z](https://doi.org/10.1007/s11136-015-1054-z)] [Medline: [26101141](https://pubmed.ncbi.nlm.nih.gov/26101141/)]
5. Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. *Proc Natl Acad Sci U S A* 2018 Jul 3;115(27):E6106-E6115 [FREE Full text] [doi: [10.1073/pnas.1711978115](https://doi.org/10.1073/pnas.1711978115)] [Medline: [29915059](https://pubmed.ncbi.nlm.nih.gov/29915059/)]
6. Hamaker EL, Wichers M. No time like the present. *Curr Dir Psychol Sci* 2017 Feb 8;26(1):10-15. [doi: [10.1177/0963721416666518](https://doi.org/10.1177/0963721416666518)]
7. 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. *J Abnorm Psychol* 2008 Aug;117(3):647-661. [doi: [10.1037/a0012532](https://doi.org/10.1037/a0012532)] [Medline: [18729616](https://pubmed.ncbi.nlm.nih.gov/18729616/)]
8. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. *Psychol Assess* 2009 Dec;21(4):463-475. [doi: [10.1037/a0017075](https://doi.org/10.1037/a0017075)] [Medline: [19947781](https://pubmed.ncbi.nlm.nih.gov/19947781/)]
9. Houben M, van den Noortgate W, Kuppens P. The relation between short-term emotion dynamics and psychological well-being: a meta-analysis. *Psychol Bull* 2015 Jul;141(4):901-930. [doi: [10.1037/a0038822](https://doi.org/10.1037/a0038822)] [Medline: [25822133](https://pubmed.ncbi.nlm.nih.gov/25822133/)]
10. Kuppens P, Allen NB, Sheeber LB. Emotional inertia and psychological maladjustment. *Psychol Sci* 2010 Jul;21(7):984-991 [FREE Full text] [doi: [10.1177/0956797610372634](https://doi.org/10.1177/0956797610372634)] [Medline: [20501521](https://pubmed.ncbi.nlm.nih.gov/20501521/)]
11. Kuppens P, Verduyn P. Emotion dynamics. *Curr Opin Psychol* 2017 Oct;17:22-26. [doi: [10.1016/j.copsyc.2017.06.004](https://doi.org/10.1016/j.copsyc.2017.06.004)] [Medline: [28950968](https://pubmed.ncbi.nlm.nih.gov/28950968/)]

12. Hershfield HE, Scheibe S, Sims TL, Carstensen LL. When feeling bad can be good: mixed emotions benefit physical health across adulthood. *Soc Psychol Personal Sci* 2013 Jan;4(1):54-61 [FREE Full text] [doi: [10.1177/1948550612444616](https://doi.org/10.1177/1948550612444616)] [Medline: [24032072](https://pubmed.ncbi.nlm.nih.gov/24032072/)]
13. Schneider S, Stone AA. Mixed emotions across the adult life span in the United States. *Psychol Aging* 2015 Jun;30(2):369-382 [FREE Full text] [doi: [10.1037/pag0000018](https://doi.org/10.1037/pag0000018)] [Medline: [25894487](https://pubmed.ncbi.nlm.nih.gov/25894487/)]
14. 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 Feb;6(1):139-149. [doi: [10.1037/1528-3542.6.1.139](https://doi.org/10.1037/1528-3542.6.1.139)] [Medline: [16637757](https://pubmed.ncbi.nlm.nih.gov/16637757/)]
15. Neubauer AB, Scott SB, Sliwinski MJ, Smyth JM. How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span. *J Pers Soc Psychol* 2020 Jul;119(1):185-203. [doi: [10.1037/pspp0000248](https://doi.org/10.1037/pspp0000248)] [Medline: [31070397](https://pubmed.ncbi.nlm.nih.gov/31070397/)]
16. Schneider S, Stone AA, Schwartz JE, Broderick JE. Peak and end effects in patients' daily recall of pain and fatigue: a within-subjects analysis. *J Pain* 2011 Feb;12(2):228-235 [FREE Full text] [doi: [10.1016/j.jpain.2010.07.001](https://doi.org/10.1016/j.jpain.2010.07.001)] [Medline: [20817615](https://pubmed.ncbi.nlm.nih.gov/20817615/)]
17. Du H, Wang L. Reliabilities of intraindividual variability indicators with autocorrelated longitudinal data: implications for longitudinal study designs. *Multivariate Behav Res* 2018;53(4):502-520. [doi: [10.1080/00273171.2018.1457939](https://doi.org/10.1080/00273171.2018.1457939)] [Medline: [29683725](https://pubmed.ncbi.nlm.nih.gov/29683725/)]
18. Flint RF. An introduction to the understanding america study internet panel. *Soc Sec Bull* 1949;60(9):1305. [doi: [10.1130/0016-7606\(1949\)60\[1305:i\]2.0.co;2](https://doi.org/10.1130/0016-7606(1949)60[1305:i]2.0.co;2)]
19. Kapteyn A. Measurement and experimentation in the social sciences. SSRN J 2010:-. [doi: [10.2139/ssrn.1889312](https://doi.org/10.2139/ssrn.1889312)]
20. Ono M, Schneider S, Junghaenel DU, Stone AA. What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. *J Med Internet Res* 2019 Feb 5;21(2):e11398 [FREE Full text] [doi: [10.2196/11398](https://doi.org/10.2196/11398)] [Medline: [30720437](https://pubmed.ncbi.nlm.nih.gov/30720437/)]
21. Dejonckheere E, Mestdagh M, Houben M, Rutten I, Sels L, Kuppens P, et al. Complex affect dynamics add limited information to the prediction of psychological well-being. *Nat Hum Behav* 2019 May;3(5):478-491. [doi: [10.1038/s41562-019-0555-0](https://doi.org/10.1038/s41562-019-0555-0)] [Medline: [30988484](https://pubmed.ncbi.nlm.nih.gov/30988484/)]
22. Eid M, Diener E. Intraindividual variability in affect: reliability, validity, and personality correlates. *J Pers Soc Psychol* 1999 Apr;76(4):662-676. [doi: [10.1037/0022-3514.76.4.662](https://doi.org/10.1037/0022-3514.76.4.662)]
23. Bringmann LF, Vissers N, Wichers M, Geschwind N, Kuppens P, Peeters F, et al. A network approach to psychopathology: new insights into clinical longitudinal data. *PLoS One* 2013;8(4):e60188 [FREE Full text] [doi: [10.1371/journal.pone.0060188](https://doi.org/10.1371/journal.pone.0060188)] [Medline: [23593171](https://pubmed.ncbi.nlm.nih.gov/23593171/)]
24. Lydon-Staley DM, Xia M, Mak HW, Fosco GM. Adolescent emotion network dynamics in daily life and implications for depression. *J Abnorm Child Psychol* 2019 Apr;47(4):717-729 [FREE Full text] [doi: [10.1007/s10802-018-0474-y](https://doi.org/10.1007/s10802-018-0474-y)] [Medline: [30203118](https://pubmed.ncbi.nlm.nih.gov/30203118/)]
25. Schimmack U. Pleasure, displeasure, and mixed feelings: are semantic opposites mutually exclusive? *Cognit Emot* 2001 Jan;15(1):81-97. [doi: [10.1080/02699930126097](https://doi.org/10.1080/02699930126097)]
26. Larsen JT, Hershfield HE, Stastny BJ, Hester N. On the relationship between positive and negative affect: their correlation and their co-occurrence. *Emotion* 2017 Mar;17(2):323-336. [doi: [10.1037/emo0000231](https://doi.org/10.1037/emo0000231)] [Medline: [27709977](https://pubmed.ncbi.nlm.nih.gov/27709977/)]
27. Kaplan KJ. On the ambivalence-indifference problem in attitude theory and measurement: a suggested modification of the semantic differential technique. *Psychol Bull* 1972;77(5):361-372. [doi: [10.1037/h0032590](https://doi.org/10.1037/h0032590)]
28. Priester JR, Petty RE. The gradual threshold model of ambivalence: relating the positive and negative bases of attitudes to subjective ambivalence. *J Pers Soc Psychol* 1996;71(3):431-449. [doi: [10.1037/0022-3514.71.3.431](https://doi.org/10.1037/0022-3514.71.3.431)] [Medline: [8831157](https://pubmed.ncbi.nlm.nih.gov/8831157/)]
29. Dejonckheere E, Mestdagh M, Houben M, Erbas Y, Pe M, Koval P, et al. The bipolarity of affect and depressive symptoms. *J Pers Soc Psychol* 2018 Feb;114(2):323-341. [doi: [10.1037/pspp0000186](https://doi.org/10.1037/pspp0000186)] [Medline: [29389216](https://pubmed.ncbi.nlm.nih.gov/29389216/)]
30. Lindquist K, Barrett L. Emotional complexity. In: Lewis M, Haviland-Jones JM, Barrett LF, editors. *Handbook of Emotions*. Third Edition. New York, USA: Guilford Press; 2008:513-530.
31. Singer J, Willett J. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford, UK: Oxford University Press; 2003.
32. Ware J. *SF-36 Health Survey: Manual & Interpretation Guide*. Boston, UK: The Health Institute, New England Medical Center; 1993.
33. Mason R, Lind D, Marchal W. *Statistics: An Introduction*. New York, USA: Harcourt Brace Jovanovich Inc; 1983.
34. Muthén L, Muthén B. *Mplus: Statistical Analysis with Latent Variables: User's Guide (Version 8)*. Los Angeles, CA: Muthén & Muthén; 2017.
35. Epskamp S, Cramer AO, Waldorp LJ, Schmittmann VD, Borsboom D. Network visualizations of relationships in psychometric data. *J Stat Soft* 2012;48(4):- [doi: [10.18637/jss.v048.i04](https://doi.org/10.18637/jss.v048.i04)]
36. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. *Proc Natl Acad Sci U S A* 2004 Mar 16;101(11):3747-3752 [FREE Full text] [doi: [10.1073/pnas.0400087101](https://doi.org/10.1073/pnas.0400087101)] [Medline: [15007165](https://pubmed.ncbi.nlm.nih.gov/15007165/)]

37. Larsen J, Hemenover S, Norris C, Cacioppo J. Turning adversity to advantage: on the virtues of the coactivation of positive-negative emotions. In: Aspinwall LG, Staudinger UM, editors. *A Psychology of Human Strengths: Fundamental Questions and Future*. Washington, DC: American Psychological Association; 2003.
38. Broderick JE, Schwartz JE, Schneider S, Stone AA. Can end-of-day reports replace momentary assessment of pain and fatigue? *J Pain* 2009 Mar;10(3):274-281 [FREE Full text] [doi: [10.1016/j.jpain.2008.09.003](https://doi.org/10.1016/j.jpain.2008.09.003)] [Medline: [19070550](https://pubmed.ncbi.nlm.nih.gov/19070550/)]
39. Diener E, Larsen RJ, Levine S, Emmons RA. Intensity and frequency: dimensions underlying positive and negative affect. *J Pers Soc Psychol* 1985;48(5):1253-1265. [doi: [10.1037/0022-3514.48.5.1253](https://doi.org/10.1037/0022-3514.48.5.1253)]
40. Dockray S, Grant N, Stone AA, Kahneman D, Wardle J, Steptoe A. A comparison of affect ratings obtained with ecological momentary assessment and the day reconstruction method. *Soc Indic Res* 2010 Nov;99(2):269-283 [FREE Full text] [doi: [10.1007/s11205-010-9578-7](https://doi.org/10.1007/s11205-010-9578-7)] [Medline: [21113328](https://pubmed.ncbi.nlm.nih.gov/21113328/)]
41. Kim J, Kikuchi H, Yamamoto Y. Systematic comparison between ecological momentary assessment and day reconstruction method for fatigue and mood states in healthy adults. *Br J Health Psychol* 2013 Feb;18(1):155-167. [doi: [10.1111/bjhp.12000](https://doi.org/10.1111/bjhp.12000)] [Medline: [23017062](https://pubmed.ncbi.nlm.nih.gov/23017062/)]
42. Han W, Feng X, Zhang M, Peng K, Zhang D. Mood states and everyday creativity: employing an experience sampling method and a day reconstruction method. *Front Psychol* 2019;10:1698 [FREE Full text] [doi: [10.3389/fpsyg.2019.01698](https://doi.org/10.3389/fpsyg.2019.01698)] [Medline: [31379699](https://pubmed.ncbi.nlm.nih.gov/31379699/)]
43. Christodoulou C, Schneider S, Stone AA. Validation of a brief yesterday measure of hedonic well-being and daily activities: comparison with the day reconstruction method. *Soc Indic Res* 2014 Feb 1;115(3):907-917 [FREE Full text] [doi: [10.1007/s11205-013-0240-z](https://doi.org/10.1007/s11205-013-0240-z)] [Medline: [24683288](https://pubmed.ncbi.nlm.nih.gov/24683288/)]
44. Miron-Shatz T, Stone A, Kahneman D. Memories of yesterday's emotions: does the valence of experience affect the memory-experience gap? *Emotion* 2009 Dec;9(6):885-891. [doi: [10.1037/a0017823](https://doi.org/10.1037/a0017823)] [Medline: [20001131](https://pubmed.ncbi.nlm.nih.gov/20001131/)]
45. Steptoe A, Wardle J. Positive affect measured using ecological momentary assessment and survival in older men and women. *Proc Natl Acad Sci U S A* 2011 Nov 8;108(45):18244-18248 [FREE Full text] [doi: [10.1073/pnas.1110892108](https://doi.org/10.1073/pnas.1110892108)] [Medline: [22042845](https://pubmed.ncbi.nlm.nih.gov/22042845/)]
46. Conner TS, Barrett LF. Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. *Psychosom Med* 2012 May;74(4):327-337 [FREE Full text] [doi: [10.1097/PSY.0b013e3182546f18](https://doi.org/10.1097/PSY.0b013e3182546f18)] [Medline: [22582330](https://pubmed.ncbi.nlm.nih.gov/22582330/)]
47. Pressman SD, Jenkins BN, Moskowitz JT. Positive affect and health: what do we know and where next should we go? *Annu Rev Psychol* 2019 Jan 4;70:627-650. [doi: [10.1146/annurev-psych-010418-102955](https://doi.org/10.1146/annurev-psych-010418-102955)] [Medline: [30260746](https://pubmed.ncbi.nlm.nih.gov/30260746/)]
48. Cain AE, Depp CA, Jeste DV. Ecological momentary assessment in aging research: a critical review. *J Psychiatr Res* 2009 Jul;43(11):987-996 [FREE Full text] [doi: [10.1016/j.jpsychires.2009.01.014](https://doi.org/10.1016/j.jpsychires.2009.01.014)] [Medline: [19272611](https://pubmed.ncbi.nlm.nih.gov/19272611/)]
49. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain* 2003 Jul;104(1-2):343-351. [doi: [10.1016/s0304-3959\(03\)00040-x](https://doi.org/10.1016/s0304-3959(03)00040-x)] [Medline: [12855344](https://pubmed.ncbi.nlm.nih.gov/12855344/)]
50. Kashdan TB, Barrett LF, McKnight PE. Unpacking emotion differentiation. *Curr Dir Psychol Sci* 2015 Feb 18;24(1):10-16. [doi: [10.1177/0963721414550708](https://doi.org/10.1177/0963721414550708)]
51. Quoidbach J, Gruber J, Mikolajczak M, Kogan A, Kotsou I, Norton MI. Emotion diversity and the emotional ecosystem. *J Exp Psychol Gen* 2014 Dec;143(6):2057-2066. [doi: [10.1037/a0038025](https://doi.org/10.1037/a0038025)] [Medline: [25285428](https://pubmed.ncbi.nlm.nih.gov/25285428/)]

## Abbreviations

- ANOVA:** analysis of variance
- DRM:** day reconstruction method
- EMA:** ecological momentary assessment
- EOD:** end-of-day
- ILA:** intensive longitudinal assessment
- NA:** negative affect
- PA:** positive affect
- UAS:** Understanding America Study
- USC:** University of Southern California

*Edited by G Eysenbach; submitted 09.04.20; peer-reviewed by M Houben, E Granholm, C Simons; comments to author 04.07.20; revised version received 15.08.20; accepted 01.09.20; published 24.09.20*

*Please cite as:*

*Schneider S, Junghaenel DU, Gutsche T, Mak HW, Stone AA*

*Comparability of Emotion Dynamics Derived From Ecological Momentary Assessments, Daily Diaries, and the Day Reconstruction Method: Observational Study*

*J Med Internet Res 2020;22(9):e19201*

*URL: <http://www.jmir.org/2020/9/e19201/>*

*doi: [10.2196/19201](https://doi.org/10.2196/19201)*

*PMID: [32969835](https://pubmed.ncbi.nlm.nih.gov/32969835/)*

©Stefan Schneider, Doerte U Junghaenel, Tania Gutsche, Hio Wa Mak, Arthur A Stone. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 24.09.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.