Journal for Person-Oriented Research
2019; 5(2): 53-64
Published by the *Scandinavian Society for Person-Oriented Research*
Freely available at https://www.person-research.org
https://doi.org/10.17505/jpor.2019.06

Evidence for the Feasibility of Person-Specific Ecological Momentary Assessment Across Diverse Populations and Study Designs

Peter D. Soyster, Hannah G. Bosley, Jonathan W. Reeves, Allison D. Altman, and Aaron J. Fisher

Corresponding author:
Peter D. Soyster, Department of Psychology, 2121 Berkeley Way, University of California, Berkeley, CA, USA, 94720.
Email: petersoyster@berkeley.edu.
Telephone: 510-664-4705

To cite this article:
Soyster, P. D., Bosley, H. G., Reeves, J. W., Altman, A. D., & Fisher, A. J. (2019), Evidence for the feasibility of person-specific ecological momentary assessment across diverse populations and study designs. *Journal for Person-Oriented Research, 5*(2), 53-64.
https://doi.org/10.17505/jpor.2019.06

Abstract: Clinical psychological science has seen an exciting shift toward the use of person-specific (idiographic) approaches to studying psychopathology and change in treatment at the level of the individual. One commonly used method in idiographic research is ecological momentary assessment (EMA). EMA offers a way to sample individuals intensively – often multiple times per day – as they go about their lives. While these methods offer benefits such as greater ecological validity and streamlined data collection, many share concerns about their feasibility across diverse clinical populations. To investigate the feasibility of using EMA to study psychological processes idiographically both in- and out of the context of therapy, the present study aggregated participants across seven studies spanning diverse clinical and community populations (N = 496), all of which utilized an idiographic EMA approach to study symptoms of psychopathology (e.g., PTSD, mood and anxiety, substance abuse). In a series of linear regression models, participant and study design characteristics were used to predict compliance with EMA surveys. Across study designs, we found that (1) participants were willing to report on symptoms and mechanisms relating to a wide range of psychopathological domains; (2) on average, participants completed 82.21% (SD = 16.34%) of all EMA surveys; and (3) compliance with EMA surveys was not significantly related to participant demographics, psychological diagnosis, personality characteristics, or most study characteristics (e.g., number of surveys per day). These findings suggest feasibility of idiographic EMA for collecting the data needed to understand psychopathology and change in treatment at the level of the individual.

Keywords: Ecological momentary assessment, idiographic, feasibility

Introduction

Researchers and practitioners of psychotherapy aspire to the same goal: helping individuals in distress to understand and ameliorate their mental health problems. Both are motivated to understand the etiological and maintaining factors of psychopathology, and the mechanisms that reveal why clients benefit from therapy. Ideally, clinical research and practice should complement one another in answering these questions. However, in practice, researchers and clinicians often struggle to communicate effectively (Teachman et al., 2012), potentially limiting progress in improving psychotherapy outcomes. Traditionally, research and clinical communities have sought the same ends (understanding and supporting therapeutic change) via different means. Importantly, researchers often focus on groups of patients, relying on aggregation and statistical analysis to form a general or typical understanding of clinical phenomena (e.g., what types of intervention strategies, symptoms, and mechanism are relevant for the average individual with depression). Conversely, as a matter of course, clinicians focus on particular patients, person by person (e.g., what will be most effective for this particular depressed individual).

The aggregated (nomothetic) approach characteristic of the research world and the individualized (idiographic) lens

inherent to clinical work each offer value and potential benefits toward the goal of improving psychotherapy outcomes. While impressive gains have been made in both realms (Hoffman, Asnaani, Vonk, Sawyer, & Fang, 2012; Barlow & Nock, 2009), more will be accomplished to the extent that researchers and clinicians can work together, bridging the nomothetic and idiographic levels of analysis. One way that researchers can bridge the gap between these communities is to conduct research that is directly relevant and immediately accessible to practicing clinicians. Person-specific research methodologies—those that can be applied to individuals on a person-by-person basis—are one approach that may fit these criteria, utilizing data from individuals' everyday lives and generating insights that are directly applicable to the individual in question. To this end, we encourage psychotherapy researchers to use methods of data collection and analysis that are actionable and prescriptive, providing clinicians with clinically-relevant information.

Ecological momentary assessment (EMA) – the repeated sampling of behavioral data in daily life – has become a prominent and ubiquitous tool for measuring research participants in situ (Myin-Germeys et al., 2018). EMA methods hold the potential to lead to improvements in both research and clinical work, and crucially, EMA may aid in the integration of the two. EMA studies typically involve asking participants to complete repeated assessments – commonly multiple times per day, via a smartphone – for a period of several days or weeks. This method can be employed to measure a wide range of psychological constructs, including clinically-relevant phenomena like emotion regulation (Ebner-Priemer & Trull, 2009), health behaviors (e.g., Shiffman, Stone, & Hu, 2008; Soyster & Fisher, 2019), and even psychopharmacology (Moskowitz & Young, 2006).

Considering the shared need for researchers and clinicians to understand complex dynamic patterns in human experience and behavior, EMA offers myriad benefits over traditional assessment methods. These include greater ecological validity, the ability to assess psychological phenomena as they occur in real-time (rather than relying on retrospective reporting), and the collection of time-series data that enables statistical methods to uncover processes and dynamics that unfold across time within an individual.

Thus, EMA holds the potential to bridge several gaps at once, those that separate cross-sectional (and otherwise aggregated) research designs from the intraindividual dynamics they hope to understand, as well as those that lie between treatment research and treatment delivery. Applied to an individual, EMA enables the discovery of time-varying dynamics in symptoms and mechanisms that are specific to each client. Learning more about the patterns that are true for any given client may enable more effective clinical care; and in the longer-term, it may be possible to aggregate these nuanced time-series data to identify patterns that characterize classes, or subgroups within larger diagnostic populations. Ultimately, this could help researchers to be more idiographic (understanding dynamics of symptoms and experiences within one client leads to more accurate prediction of behavior that could be used clinically) and clinicians to be more systematic (collecting quantitative data to measure symptoms and mechanisms, arming them to contribute to the research literature).

Some researchers and clinicians are already thinking along these lines. Arguments can be found in the literature for emphasizing the importance of person-specific analyses to achieve a more nuanced understanding of the diversity of intra-individual patterns (e.g., Barlow & Nock, 2009). In the clinical realm, some have begun to collect quantitative data as a strategy to better understand treatment processes (Rubel, Zilcha-Mano, Giesemann, Prinz, & Lutz, 2019; Brown, Bosley, Kenyon, Chen, & Levenson, 2019). More recently, empirical and theoretical papers have discussed the merits of idiographic data collection and analysis applied to clinical work (Fisher, 2015; Piccirillo, Beck, & Rodebaugh, 2019).

Further, the direct application of such thinking to clinical cases has been demonstrated to yield desirable outcomes, helping patients in an open clinical trial of personalized therapy to achieve therapeutic gains over shorter periods of time (Fisher et al., 2019). Specifically, Fisher and colleagues used EMA to measure symptoms and mechanisms of depressive and anxious pathology four times per day for 30 days, and then provided a cognitive-behavioral intervention tailored to the person based on a quantitative case conceptualization derived from their pre-treatment EMA data. This procedure produced large treatment effects in an average of approximately 10 sessions; when compared to a standard 16-week course of manualized CBT, personalization helped patients get better faster. These findings underscore the potential benefits of using EMA to bridge the gap between empirical data and clinical care.

Given strong evidence for the benefits of EMA in clinical practice, it is worth considering potential costs or barriers to its use in practice. What may get in the way of applying EMA to clinical work? Are certain groups of people more or less likely to comply with this method? What characteristics of the study procedures (e.g., the sampling frequency and duration of the sampling period; how participants are incentivized) enhance or diminish compliance?

Among clinicians and researchers alike, a predominant concern is feasibility and potential participant burden. EMA is a relatively dense sampling approach that asks participants to regularly, often intensively, provide self-report data (i.e., several times per day). Despite evidence that participants can tolerate up to 60 surveys per day (Kuppens & Koval, 2012) there is reasonable concern that in clinical contexts, participants might struggle to comply with the demands of EMA sampling even once, twice, or four times per day. For a number of reasons, it is important that EMA participants answer as many surveys as possible and provide relatively complete data within a given sampling pe-

riod. Compliance with EMA surveys is important both for statistical analysis (e.g., missing data is problematic with many statistical approaches) and for interpretation of results (e.g., we don't always know why data were missing, and this could represent a source of bias). Furthermore, if EMA methods represent a barrier to research participation among certain demographic or diagnostic groups more than others, this may threaten the generalizability of findings obtained from these methods.

Some research groups have investigated predictors of EMA compliance and found people to be generally receptive and compliant (e.g., Rintala, et al., 2019; Myin-Germeys et al., 2018). As one well-powered example, Rintala and colleagues (2019) examined compliance with 4 to 6 days of assessment across more than 1,500 participants. They found that certain populations (females, persons with psychosis) exhibited lower compliance. Other groups such as Palmier-Claus and colleagues (2011) and Myin-Germeys and colleagues (2018) have reviewed the literature and called for increased focus on predictors of compliance in clinical samples, or with different sampling periods. However, even in light of these useful findings, the diversity of EMA study methodologies to date combined with the burgeoning popularity of this method warrant replication and extension of previous work in this area. To our knowledge, only a handful of studies have systematically examined predictors of EMA compliance to date. However, questions remain about its feasibility for most participants in clinical situations. It remains an empirical question whether certain participant features, such as demographics or individual differences in personality or psychopathology, or even features of the sampling procedure itself, would lead to differential rates of participation and compliance. If these identifiable characteristics are associated with compliance, this may mean that EMA is not equitably appropriate for all populations.

Extant work across a variety of contexts and samples have investigated compliance rates in EMA studies, including in chronic pain patients (Morren et al., 2007), alcohol and substance users (Sokolovsky et al., 2013), individuals with symptoms of psychosis (Hartley et al., 2013), and individuals carrying a primary diagnosis of PTSD with comorbid substance use (Possemato et al., 2012). While some of these studies found certain characteristics to be correlated with increased survey compliance (i.e., shorter survey length, compensating subjects, female gender, and higher age), others found no characteristics to be significantly correlated with compliance rates. Thus, no unifying demographic variables, clinical characteristics, nor facets of study designs have emerged as a common predictor of increased compliance rates in EMA studies.

The present study aimed to address this question, with the goal of providing data relevant to the evaluation of the feasibility of EMA as a tool in psychotherapy research broadly. Our laboratory has amassed over five years of experience conducting EMA data collection with clinical and community populations across varying levels and types of psychopathology. Collapsing across our data to date, we examined whether three classes of variables—demographic, psychological, or procedural—predicted compliance with EMA surveys. We ran five models in which these were tested as predictors of compliance. In these cases, a null result (that is, no difference in compliance as a function of these predictors) is ideal, as this would indicate that there are not systematic differences as a function of participant variables that hinder compliance. By identifying and understanding any potential differences, EMA researchers can augment the design of their studies a priori to address these disparities and maximize participant compliance.

Methods

Participants

The present analyses used data from N = 496 adults drawn from seven separate studies using EMA research designs. Participants were largely Asian (38%) and female (67%), with an average age of 24.30 (SD = 9.80). Of the non-Asian participants, 159 (32%) identified as white, 59 (12%) as multiracial/other, 68 (14%) as Hispanic/Latinx, and 20 (4%) as Black/African American. The modal annual income and modal education level reported by the present sample was less than $10,000 per year (range = <$10,000 – >$100,000) and some college education (range = some high school – post-graduate degree), respectively.

Procedure

All study procedures were approved by the University of California, Berkeley Committee for the Protection of Human Subjects. Data included in the present analyses were drawn from seven studies aimed at addressing different research questions. As such, these studies varied in the constructs measured, length of EMA assessment, and sampling frequency. However, included studies used a common framework for procedures surrounding project flow and administration of the EMA protocol. Each study included three parts: (a) participant recruitment and screening, (b) baseline assessment, and (c) EMA survey period. For each study, participants were required to complete at least 80% of the prompts sent over the length of the EMA assessment period to qualify for compensation in the form of course credit, cash, and/or cost-free psychological services.

Recruitment. Common inclusion criteria across these studies were fluency in the English language, being $\geqslant$18 years old, and having daily access to a web-enabled smartphone that could receive text messages. For a given study, participants were recruited either from (a) a large undergraduate research pool or (b) online/physical advertisements posted in the surrounding community. For those recruited from a university-based research pool, potential participants completed a pre-screening survey that included a brief demographics survey and self-report questionnaires.

Those recruited from the surrounding community contacted study staff by phone or email and completedd a brief phone or email screen. Participants that passed either preliminary screen were then invited by study staff to present at a university-based research laboratory to complete baseline measures and receive instruction on completing EMA surveys. Baseline measures used in the present analyses include the Depression Anxiety Stress Scales (DASS; Lovibond & Lovibond, 1995), the revised NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1995), the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), and the MacArthur Scale of Subjective Social Status (MSSS; Adler et al., 1994). For four of the seven samples, one of the following clinical interviews were used to determine psychiatric diagnosis at baseline: the Mini-International Psychiatric Interview (Sheehan et al., 1998), the Anxiety and Related Disorders Interview Schedule for DSM-5 (Brown & Barlow, 2014), or the Clinician Administered PTSD Scale (Blake et al., 2000).

EMA. Following enrollment in each study, participants' mobile phone numbers were entered into a secure web-based survey system. This system sent pings to participants' mobile phones with survey prompts several times per day, with each prompt received as a text message containing a hyperlink to a web-based survey. Each ping populated the back-end system with a time stamp whether the participant completed the survey or not. For a given study, participants were asked to complete at least 80% of the sent surveys for a minimum number of days. The number of prompts sent per day, the number and content of individual survey items, and length of the EMA assessment period varied by study. Table 1 provides information about the characteristics of each study.

Measures

DASS (Lovibond & Lovibond, 1995). The DASS is a 42-item measure designed to assess three related negative emotional states, including depression, anxiety, and stress. Each of the three subscales consists of 14 items and each item is rated on a 4-point Likert scale ranging from 0-3 with the anchors "did not apply to me at all", "applied to me to some degree or some of the time", "applied to me to a considerable degree or a good part of the time", and "applied to me very much or most of the time". Total scores for each subscale of the DASS were calculated by summing across the 14 items that made up each subscale.

MSSS (Adler et al., 1994). The MSSS is a single item measure designed to assess the common sense of social status across various indicators of socioeconomic status (e.g., education, income). In a pictorial format, the MSSS presents a "social ladder", with 10 rungs, and asks participants to select the rung on which they feel they stand. Scores (ranging from 1-10) were recorded for each participant, with high scores reflecting higher levels of subjectively rated social status.

NEO PI-R (Costa & McCrae, 1995). The NEO PI-R is a 60-item measure of the five major domains of personality, including neuroticism, extraversion, openness, agreeableness, and conscientiousness. Each item consists of a single statement. Participants are asked to rate the degree to which they agree with each item on a 5-point scale ranging from 1-5 with the anchors "strongly disagree", "disagree", "neutral", "agree", and "strongly agree". Total scores for each subscale on the NEO PI-R were calculated by summing the items that made up each subscale.

PANAS (Watson et al., 1988). The PANAS is a 20-item self-report measure consisting of two 10-item subscales designed to assess an individual's tendency to experience positive and negative affect. Each item contains words that describe different feelings and emotions and is rated on a 5-point Likert scale ranging from 1-5 with the anchors "very slightly or not at all", "a little", "moderately", "quite a bit", and "extremely". Total positive (PA) and negative affect (NA) scores were calculated by summing across the 10 items that made up each subscale.

Data preparation and analysis

We constructed a series of linear regression analyses to assess group-level differences in EMA survey completion rates as a function of participant demographics, study features (survey length, EMA assessment period length, sampling frequency, and compensation type), personality traits, trait affect, psychopathology, and social status. Survey compliance was operationalized as the total number of surveys a participant completed, divided by the total number of surveys sent. Given that the resulting distribution was highly negatively skewed (skewness = -2.07), this variable was reverse scored and subsequently log-transformed to approximate a normal distribution (M = 2.59, SD = 0.88, range = 0 - 4.58, skewness = -0.48) for use in all analyses (we refer to this transformed dependent variable as the log of the percentage of missed surveys). As noted above, personality traits and trait affect were indexed by total scores on subscales of the NEO PI-R and PANAS, respectively.

We created a series of factor variables to represent the presence of a given class of psychiatric diagnosis, including any depressive, anxious, psychotic, compulsive, trauma, or substance use disorder. For depressive, anxious, compulsive, and substance use disorders, these variables had three levels, including absence of a given class of disorder (0/2), presence of clinically-elevated symptoms using the DASS-42 without a confirmed diagnosis (1/2), and presence of a diagnosis confirmed by clinical interview (2/2). For trauma, psychotic, and 'other' disorders (e.g., hypomanic episode), these variables had two levels, indicating presence or absence of a confirmed diagnosis (0/1). Finally, we created a variable indicating the presence of any comorbid disorder confirmed by clinical interview (0/1).

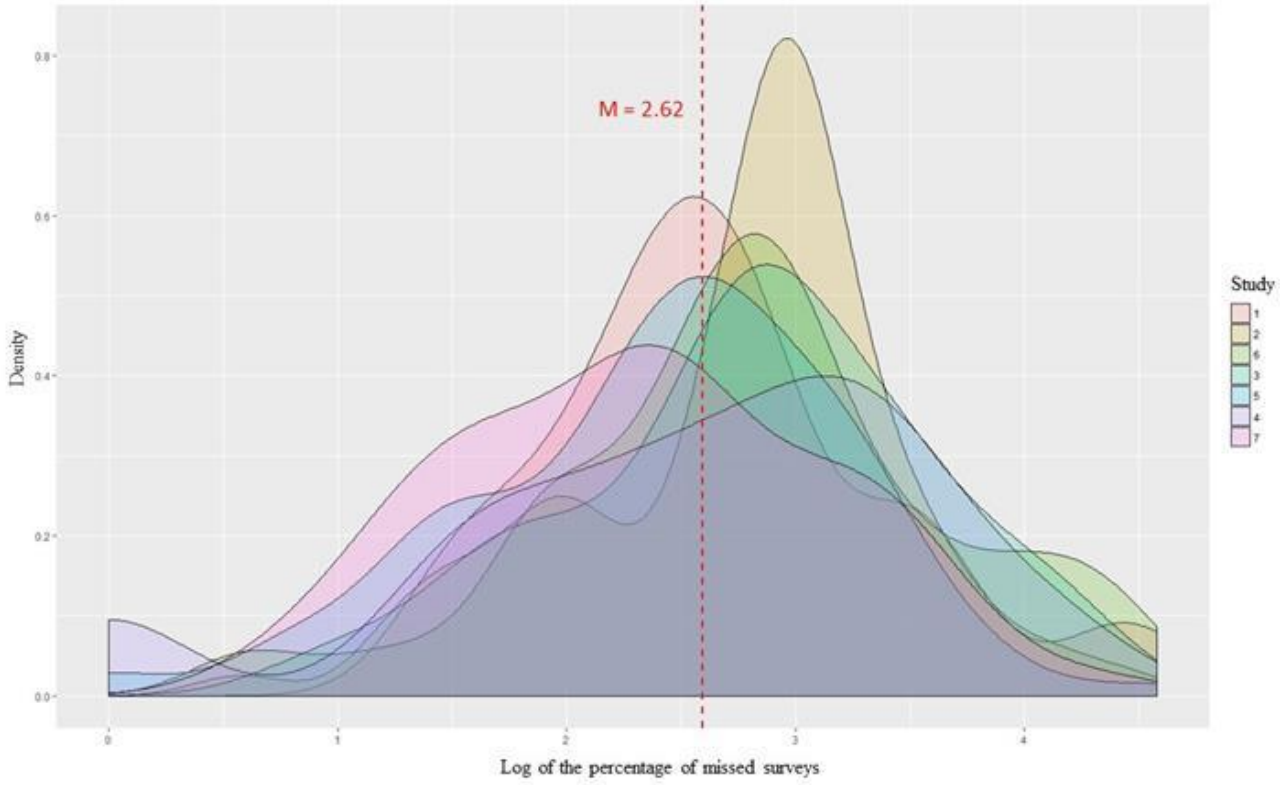

Figure 1. Density plots of the log of the percentage of missed surveys across 7 studies

Results

Full-sample analyses

Aggregating across all participants (N = 496), the average survey completion rate was 82.21% (SD = 16.34%). The difference in the percentage of completed surveys between the study with the lowest (Study 6, 77.98%) and highest (Study 7, 86.59%) completion rate was statistically significant (t = 2.73, p =.002). To address skew, the log of the percentage of missed surveys was used as the dependent variable for all regression models (See Figure 1).

Table 2 presents the results of the regression analysis using participant demographics and study characteristics to predict differences in survey compliance. In total, the model explained approximately 4% of the variance in the dependent variable ($R^2_{adjusted}$ = 0.038, p = 0.007). The model indicated that receiving combination compensation (β = 0.48, SE = 0.23, t = 2.06, p = 0.04) was associated with a higher percentage of missed surveys. None of the other included variables were significantly associated with survey compliance at the α = 0.05 level.

Sub-sample analyses

Table 3 presents the results of all subsample analyses.

Big Five Personality. N = 228 participants completed the NEO personality inventory as part of their study participation. This sample had a mean openness score of 37.39 (SD = 11.08), a mean conscientiousness score of 36.68 (SD = 10.09), a mean extraversion score of 36.07 (SD = 8.65), and a mean agreeableness score of 35.30 (SD = 10.00), and a mean neuroticism score of 37.22 (SD = 11.43). In total, the model explained 0 % of the variance in the dependent variable ($R^2_{adjusted}$ = -0.02, p = 0.96). None of the five personality factors were significantly related to the log of the percentage of missed surveys (all ps >.05).

Table 1.
Study characteristics

	Construct Measured	# Items	Sampling Frequency	Sampling Period (days)
Study 1	Alcohol Use	16	8	15
Study 2	Tobacco Use	40	4	30
Study 3	PTSD Symptoms	37	4	30
Study 4	Dampening	26	4	7
Study 5	Positive and Negative Affect	23	4	21
Study 6	MDD and GAD Symptoms	22	4	30
Study 7	Social Media Use and Mood	18	8	14

Note. # Items = total number of items per survey prompt; PTSD = posttraumatic stress disorder; MDD = major depressive disorder; GAD = generalized anxiety disorder

Table 2
Effect of demographic characteristics and EMA sampling features on EMA survey compliance

		β	***SE***	***t***	***p***
Intercept		2.91	0.42	6.87	< .001
Age		-0.19	0.11	-1.61	.11
Sex	Female	-	-	-	-
	Male	0.08	0.09	0.87	.38
Race	White/Caucasian	-	-	-	-
	Asian	0.11	0.10	1.190	.23
	Black/African American	0.36	0.22	1.65	.10
	Hispanic/Latinx	0.11	0.13	0.88	.38
	Multiracial	0.03	0.19	0.18	.86
	Other	-0.11	0.17	-0.65	.52
Education	Some high school	-	-	-	-
	High school diploma	-0.41	0.45	-0.91	.36
	Some college	-0.44	0.43	-1.02	.31
	Bachelor's degree	-0.60	0.43	-1.40	.16
	Post-graduate degree	-0.79	0.44	-1.78	.08
# Surveys Sent		0.05	0.16	0.34	.74
# Survey Items		0.14	0.15	0.95	.34
Sampling Period (Days)	7	-	-	-	-
	14	-0.11	0.26	-0.42	.68
	15	0.23	0.30	0.75	.46
	21	-0.15	0.17	-0.90	.37
	30	0.11	0.24	0.47	.64
Compensation	Course credit	-	-	-	-
	Money	0.10	0.18	0.54	.59
	Combination with services	0.48	0.23	2.07	.04

Note. # surveys sent = number of surveys over full sampling period; # survey items = number of survey items in each EMA survey

Table 3
Subsample analyses of predictors of EMA survey compliance

Big Five Personality Model (n = 228)					
		β	*SE*	*t*	*p*
Intercept		2.77	0.05	53.72	< .001
Openness		-0.02	0.14	-0.17	.86
Conscientiousness		0.07	0.14	0.48	.64
Extraversion		0.07	0.13	0.57	.57
Agreeableness		-0.05	0.15	-0.33	.74
Neuroticism		-0.01	0.13	-0.10	.92
Positive and Negative Affect Scale Model (n = 328)					
		β	*SE*	*t*	*p*
Intercept		2.58	0.05	50.50	< .001
Trait NA		0.06	0.11	0.56	.58
Trait PA		-0.02	0.11	-0.21	.84
Psychiatric Diagnosis Model (n = 239)					
		β	*SE*	*t*	*p*
Intercept		2.84	0.11	26.47	< .001
MDD	No	-	-	-	-
	Elevated	-0.27	0.18	-1.46	.15
	Yes	-0.17	0.19	-0.88	.38
Any anxiety disorder	No	-	-	-	-
	Elevated	0.23	0.20	1.12	.27
	Yes	0.12	0.20	0.62	.54
Any compulsive disorder	No	-	-	-	-
	Elevated	-0.59	0.42	-1.42	.16
	Yes	0.34	0.27	1.25	.21
PTSD	No	-	-	-	-
	Yes	0.022	0.16	0.14	.89
Any psychotic disorder	No	-	-	-	-
	Yes	-0.28	0.32	-0.87	.38
SUD	No	-	-	-	-
	Elevated	-0.31	0.30	-1.05	.30
	Yes	-0.12	0.14	-0.90	.37
Any other diagnosis	No	-	-	-	-
	Yes	0.03	0.18	0.16	.87
Any comorbid diagnosis	No	-	-	-	-
	Yes	-0.05	0.20	-0.24	.81
MacArthur Subjective Social Status (n = 363)					
		β	*SE*	*t*	*p*
Intercept		2.51	0.46	54.83	< .001
MSSS		-0.07	0.09	-0.75	.45

Note. NA = negative affect; PA = positive affect; MDD = major depressive disorder; PTSD = post-traumatic stress disorder; MSSS = MacArthur Subjective Social Status.

Table 4
Participant count by diagnostic group and study

		S1	S2	S3	S4	S5	S6	S7	Total
N		33	80	26	96	107	100	54	496
MDD	No	8	27	2	21	52	48	42	200
	Elevated	19	33	15	75	55	15	12	224
	Yes	6	20	9	-	-	37	-	72
	Not assessed	0	0	0	0	0	0	0	0
Anxiety	No	8	29	0	6	34	42	33	152
	Elevated	13	26	11	90	73	7	21	170
	Yes	12	25	15	-	-	51	-	174
	Not assessed	0	0	0	0	0	0	0	0
Compulsive disorder	No	29	70	25	-	-	100	-	224
	Elevated	3	0	1	-	-	0	-	4
	Yes	1	10	0	-	-	0	-	11
	Not assessed	0	0	0	96	107	0	54	257
PTSD	No	33	72	0	-	-	95	-	200
	Yes	0	8	26	-	-	5	-	39
	Not assessed	0	0	0	96	107	0	54	257
Psychotic disorder	No	32	73	26	-	-	100	-	231
	Yes	1	7	0	-	-	0	-	8
	Not assessed	0	0	0	96	107	0	54	257
SUD	No	17	32	22	-	-	98	-	169
	Elevated	1	7	0	-	-	0	-	8
	Yes	15	41	4	-	-	2	-	62
	Not assessed	0	0	0	96	107	0	54	257
Other diagnosis	No	29	62	25	-	-	97	-	213
	Yes	4	18	1	-	-	3	-	26
	Not assessed	0	0	0	96	107	0	54	257
Any comorbidity	No	23	49	7	-	-	65	-	251
	Yes	10	31	19	-	-	35	-	95
	Not assessed	0	0	0	96	107	0	54	150

Note. S1-S7 = Study 1 - Study 7; MDD = major depressive disorder; Anxiety = any anxiety disorder; PTSD = post-traumatic stress disorder; Psychotic disorder = any psychotic disorder; SUD = any substance use disorder.

Trait affect. *N* = 328 participants completed the PANAS as part of their study participation. The average trait negative affect was 25.17 (*SD* = 8.37) and the average trait positive affect was 28.66 (*SD* = 9.57). In total, the model explained effectively 0% of the variance in the dependent variable ($R^2_{adjusted}$ = -0.005, p = 0.80). Neither trait positive or negative affect were significantly related to the log of the percentage of missed surveys (*p*s = .58 and .84, respectively).

Psychiatric diagnosis. *N* = 239 participants were assessed for psychiatric diagnosis as part of their study participation. The frequency of each diagnostic category is presented in Table 4. In total, the model explained effectively 0% of the variance in the dependent variable ($R^2_{adjusted}$ = -0.009, p = 0.62). There were no significant differences in the log of the percentage of missed surveys among the psychiatric diagnosis categories (all *p*s >.05).

Subjective social status. *N* = 363 participants completed the MSSS as part of their participation. The average social status rating was 6.31 (*SD* = 1.84). Subjective social status was not found to be related to survey compliance ($R^2_{adjusted}$ = -0.001, p = 0.45).

Discussion

The field of psychotherapy research has seen a recent increase in the use of momentary assessment methods such as EMA to understand mechanisms of pathology and therapeutic change processes. EMA offers many potential benefits, including a more nuanced idiographic understanding of clients' experiences as they occur in their ecological contexts, improved clinical care, and greater integration of clinical science and practice – but some concerns about broad feasibility of EMA stand in the way of these potential benefits coming to fruition.

In the present study, we sought to evaluate a common,

intuitive concern about the feasibility of EMA: specifically, there is a concern that certain groups of people might systematically exhibit lower compliance with EMA methods relative to others (Palmier-Claus et al., 2011). Therefore, we aimed to establish whether characteristics of participants (such as demographics, personality, and psychopathology) or of the EMA sampling procedure (e.g., frequency of surveys per day, duration of sampling period) were significantly associated with EMA compliance. Across data from seven EMA studies, which represented a diversity of participant populations and sampling procedures, we found no evidence for associations between any group-level participant characteristic and EMA compliance. With the exception of one finding (discussed below), we also did not find support for characteristics of the sampling procedure as predictors of EMA compliance. As previous research has demonstrated overall compliance rates ranging from 33% (Courvouiser et al., 2012) to 90% (Sokolovsky et al., 2013), the overall rates from the present study (82.21%) is on the higher end of that range.

Broadly, our set of null findings here suggests that – while there are certainly individual differences in compliance with EMA studies – group-level characteristics such as demographics, personality, psychopathology, or how people are sampled do not seem to systematically affect EMA compliance. Based on these empirical data, we conclude that there is not strong evidence for the belief that certain groups of people would be significantly more or less compliant with EMA procedures.

The present study builds upon extant literature in a number of important ways. First, many prior studies such as Rintala et al. (2019) utilized a paper-and-pencil assessment procedure, whereas the present study conducted sampling via smartphone. As smartphones become more readily accessible, future studies may shift to this method, so it is important to glean a better understanding of factors that affect this type of sampling procedure. Further, smartphone sampling as conducted in the present studies enables an objective, time-stamped measure of compliance – in paper-and-pencil studies, as Rintala and colleagues point out, it is possible that participants could fill out the diary assessments retrospectively without investigators' knowledge, thereby biasing/altering estimates of compliance. A second, novel extension offered by the present findings is that the seven studies analyzed here utilized longer sampling periods than have been previously reported in the literature (ranging from 7 to 30 days across our datasets). In general, the previous literature on EMA compliance has investigated studies with much shorter sampling periods (Rintala et al., 2019). It is important to examine compliance over longer EMA sampling periods to determine whether the increased burden leads to fatigue or drop-off in compliance over the longer interval. A longer sampling period is often useful to better understand time-varying dynamics in psychopathology and psychotherapy, as certain patterns (e.g., change in therapy) may unfold over a longer time scale than could be potentially captured with a period of only a few days.

In the literature, compliance estimates range from ~70% to ~80% across empirical studies, reviews, and meta-analyses (Hartley et al., 2014; Jones et al., 2018; Myin-Germeys et al., 2018; Palmer-Claus et al., 2011; Rintala et al., 2019;). Generally, our data are concordant with estimates of these other published studies, with an average compliance of 82.21% ranging from 77.98% to 86.59%. This suggests that the use of smartphone assessment can be expected to yield compliance that is on par with other previously studied methods such as paper-and-pencil diaries or wrist-watch and palm-pilot studies. Further, the concordance between our findings and estimates in the literature (as well as comparisons available within the present study, from 7-days to 30-days) also suggests that the longer sampling period would not necessarily lead to reductions in compliance.

There have been mixed findings in the literature regarding predictors of compliance, which may reflect the diverse methodologies that can be employed in an EMA study. While some prior work has found gender, psychopathology (Rintala et al., 2019), or details of the sampling procedure (Jones et al., 2018) to predict EMA compliance rates, the present study did not replicate these findings. With one exception, we did not find significant associations between any of our tested predictor variables and EMA compliance. Notably, we did find that the manner of compensation for study participation predicted compliance (specifically, receiving a combination of therapy services and either money or course credit predicted slightly lower rates of compliance with EMA surveys). However, even the sample in our study with the lowest compliance (78%) is equivalent to average estimates in the literature (Myin-Germeys et al., 2018). Thus, we believe that this effect should be interpreted with some caution, and warrants further exploration and replication.

One possible explanation for this effect is the presence of slight systematic differences in our samples. That is, the one open-trial which utilized a combination of payment and therapy services to incentivize compliance recruited community participants, many of whom were experiencing mental-health problems and seeking treatment. Comparing this population to a higher-functioning undergraduate sample (which would be incentivized with course credit alone) may yield slight differences. Another possible explanation for this effect draws from the literature on motivation and incentive theory. Specifically, in light of findings that providing external incentives (such as money or course credit) may reduce intrinsic motivation to complete a task (Wiersma, 1992), perhaps a combination of incentives led to decreased motivation and decreased task performance. Future studies should seek to determine how study compensation may affect compliance with EMA; tailoring compensation to optimally incentivize compliance would be a worthwhile endeavor for future research and

practice.

Limitations and Future Directions

While the present study represents a novel extension of the literature on EMA compliance to date, it is not without limitations. Despite the fact that the sample size for the present analysis was large, participants were from a relatively circumscribed geographic area (the San Francisco Bay Area), a part of the country where technology and cell phone usage is quite high. Future research should investigate compliance with a nationwide sample to investigate the generalizability of these findings to broader locations. Furthermore, some of the studies included in our analyses had methods for checking in on and reminding participants to respond to the surveys, yet this was not accounted for in our method. Future research should code for type, frequency, and consistency of reminders to see if that may influence compliance rates. Additionally, while we have assessed compliance in this study, we have not assessed other strict metrics of data quality; while compliance is important for EMA research, high compliance does not necessarily equate to high quality data; future work should investigate data quality as well as compliance.

While our results suggest general feasibility of EMA methods within typical design parameters, we do not intend to suggest that such method are without upper limits. The sampling periods studied ranged from 7 to 30 days, the sampling frequencies ranged between 4 and 8 times per day, and the survey items ranged from 16 to 40. It is logical to assume that there are sampling paradigms that would be too intensive for participants to comply with. Further, while the present study involved participants with a relatively broad range of psychopathologies, the sample was restricted in ways that limit the generalizability of our findings. For example, the present sample did not include anyone who was floridly psychotic. Where are the limitations to EMA methodology? Future research should seek to understand the upper limits of EMA feasibility in relation to the frequency, duration, and intensity of the sampling paradigm.

Taken together with the other recent literature on this topic, we hope the present findings will serve to alleviate concerns over the feasibility of EMA methods, leading to the broader application of such methods to study psychopathology and therapeutic change in diverse populations. In the absence of other evidence, we suggest that it is wise to assume, regardless of their group characteristics, that people are generally able to comply with EMA. As EMA methods expand and are utilized more broadly, future studies should continue to investigate individual differences in compliance, with an eye toward optimizing/tailoring the procedure to be amenable to different individual needs.

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

Author contributions

PS conceived the present study, ran all statistical analyses, and wrote the manuscript with support from HB, JR, AA, and AF. All authors provided data for analyses and contributed to discussion of results and formalizing the final manuscript.

Action editor

Lars-Gunnar Lundh served as action editor for this article.

References

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., & Syme, S. L. (1994). Socioeconomic status and health. The challenge of the gradient. *The American Psychologist, 49*(1), 15–24. https://doi.org/10.1037/0003-066X.49.1.15

Barlow, D. H., & Nock, M. K. (2009). Why can't we be more idiographic in our research? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 4*(1), 19–21. https://doi.org/10.1111/j.1745-6924.2009.01088.x

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D., Klauminzer, G., Charney, D. S., … Buckley, T. C. (2000). Clinician-Administered PTSD Scale (CAPS): Instruction Manual. Boston, MA: National Center for PTSD.

Brown, T. A., & Barlow, D. H. (2014). *Anxiety and Related Disorders Interview Schedule for DSM-5.* New York, NY: Oxford University Press.

Brown, C. L., Bosley, H. G., Kenyon, A. D., Chen, K.-H., & Levenson, R. W. (2019). An idiographic statistical approach to clinical hypothesis testing for routine psychotherapy: A case study. *Behaviour Research and Therapy, 118*, 43–53. https://doi.org/10.1016/j.brat.2019.03.014

Ebner-Priemer, U. W., & Trull, T. J. (2009). Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological Assessment, 21*(4), 463–475. https://doi.org/10.1037/a0017075

Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: Implications for personalized care. *Journal of Consulting and Clinical Psychology, 83*(4), 825–836. https://doi.org/10.1037/ccp0000026

Fisher, A. J., Bosley, H. G., Fernandez, K. C., Reeves, J. W., Soyster, P. D., Diamond, A. E., & Barkin, J. (2019). Open trial of a personalized modular treatment for mood and anxiety. *Behaviour Research and Therapy, 116*, 69–79. https://doi.org/10.1016/j.brat.2019.01.010

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences, 115*(27), E6106. https://doi.org/10.1073/pnas.1711978115

Granholm, E., Loh, C., & Swendsen, J. (2008). Feasibility

and Validity of Computerized Ecological Momentary Assessment in Schizophrenia. *Schizophrenia Bulletin, 34*(3), 507–514. https://doi.org/10.1093/schbul/sbm113

Hartley, S., Varese, F., Vasconcelos e Sa, D., Udachina, A., Barrowclough, C., Bentall, R. P.,… Palmier-Claus, J. (2014). Compliance in experience sampling methodology: The role of demographic and clinical characteristics. Psychosis: *Psychological, Social and Integrative Approaches, 6*(1), 70–73. https://doi.org/10.1080/17522439.2012.752520

Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. *Cognitive Therapy and Research, 36*(5), 427–440. https://doi.org/10.1007/s10608-012-9476-1

Jones, A., Remmerswaal, D., Verveer, I., Robinson, E., Franken, I. H. A., Wen, C. K. F., & Field, M. (2019). Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. *Addiction, 114*(4), 609–619. https://doi.org/10.1111/add.14503

Costa Jr, P. T. C., & McCrae, R. R. (1995). Domains and Facets: Hierarchical Personality Assessment Using the Revised NEO Personality Inventory. *Journal of Personality Assessment, 64*(1), 21–50. https://doi.org/10.1207/s15327752jpa6401_2

Koval, P., & Kuppens, P. (2012). Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion, 12*(2), 256–267. https://doi.org/10.1037/a0024756

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behavior Research and Therapy, 33*(3), 335-343. https://doi.org/10.1016/0005-7967(94)00075-u

Morren, M., Dulmen, S. van, Ouwerkerk, J., & Bensing, J. (2009). Compliance with momentary pain measurement using electronic diaries: A systematic review. *European Journal of Pain, 13*(4), 354–365. https://doi.org/10.1016/j.ejpain.2008.05.010

Moskowitz, D. S., & Young, S. N. (2006). Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. *Journal of Psychiatry & Neuroscience: JPN, 31*(1), 13–20.

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry, 17*(2), 123–132. https://doi.org/10.1002/wps.20513

Palmier-Claus, J. E., Myin-Germeys, I., Barkus, E., Bentley, L., Udachina, A., Delespaul, P. A. E. G.,… Dunn, G. (2011). Experience sampling research in individuals with mental illness: reflections and guidance. *Acta Psychiatrica Scandinavica, 123*(1), 12–20. https://doi.org/10.1111/j.1600-0447.2010.01596.x

Piccirillo, M. L., Beck, E. D., & Rodebaugh, T. L. (2019). A Clinician's Primer for Idiographic Research: Considerations and Recommendations. *Behavior Therapy*, S000578941930022X. https://doi.org/10.1016/j.beth.2019.02.002

Possemato, K., Kaier, E., Wade, M., Lantinga, L. J., Maisto, S. A., & Ouimette, P. (2012). Assessing daily fluctuations in posttraumatic stress disorder symptoms and substance use with interactive voice response technology: Protocol compliance and reactions. *Psychological Services, 9*(2), 185–196. https://doi.org/10.1037/a0027144

Rintala, A., Link to external site, this link will open in a new window, Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2019). Response compliance and predictors thereof in studies using the experience sampling method. *Psychological Assessment, 31*(2), 226–235. http://dx.doi.org/10.1037/pas0000662

Rubel, J. A., Zilcha-Mano, S., Giesemann, J., Prinz, J., & Lutz, W. (2019). Predicting personalized process-outcome associations in psychotherapy using machine learning approaches. A demonstration. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 1–10. https://doi.org/10.1080/10503307.2019.1597994

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., … Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry, 59* Suppl 20, 22-33;quiz 34-57.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Sokolovsky, A. W., Mermelstein, R. J., & Hedeker, D. (2014). Factors predicting compliance to ecological momentary assessment among adolescent smokers. *Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco, 16*(3), 351–358. https://doi.org/10.1093/ntr/ntt154

Soyster, P. D., & Fisher, A. J. (2019). Involving stakeholders in the design of ecological momentary assessment research: An example from smoking cessation. *PLOS ONE, 14*(5), e0217150. https://doi.org/10.1371/journal.pone.0217150

Teachman, B. A., Drabick, D. A. G., Hershenberg, R., Vivian, D., Wolfe, B. E., & Goldfried, M. R. (2012). Bridging the gap between clinical research and clinical practice: Introduction to the special section. *Psychotherapy, 49*(2), 97–100. https://doi.org/10.1037/a0027346

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070.

https://doi.org/10.1037/0022-3514.54.6.1063
Wiersma, U. J. (1992). The effects of extrinsic rewards in intrinsic motivation: A meta-analysis. *Journal of Occupational and Organizational Psychology, 65*(2), 101–114. https://doi.org/10.1111/j.2044-8325.1992.tb00488.x