Current Psychology (2023) 42:6287–6301
https://doi.org/10.1007/s12144-021-01974-3

Perceived disturbance and predictors thereof in studies using the experience sampling method

Aki Rintala [1,2] · Martien Wampers [1,3] · Ginette Lafit [1,4] · Inez Myin-Germeys [1] · Wolfgang Viechtbauer [1,5]

Accepted: 7 June 2021 / Published online: 12 June 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Intensive repeated momentary diary techniques, such as the experience sampling method (ESM), place considerable burden on study participants by repeatedly prompting them to fill in a questionnaire several times throughout the day. The objective of this study was to investigate the level of disturbance reported by participants and predictors thereof using a pooled dataset of seven ESM studies including 1319 healthy participants and individuals with different mental health conditions. All studies used a design of 10 prompts per day over the course of 4 to 6 days. Participants reported no disturbance at 37% of the measurement occasions and the overall level of disturbance was relatively low. Also, clinical samples did not report significantly different levels of disturbance compared to the healthy participants. However, higher levels of disturbance were found for later study days, in the morning, and on weekends. Also, younger participants reported feeling more disturbed by the prompts. In terms of momentary states, higher levels of activity, activity stress, and tiredness were associated with higher disturbance. Also, participants experienced more disturbance when in company compared to when alone, and the level of disturbance increased when in company with less familiar people. Our findings suggest that studies using high-frequency ESM protocols can be used in mental health research, but researchers should be aware of the possible disturbance caused by their research design.

Keywords Disturbance · Burden · Experience sampling method · Data quality · Ecological momentary assessment

Introduction

Momentary assessment techniques such as the experience sampling method (ESM; also known as ecological momentary assessment) are structured paper-and-pencil or electronic diary techniques to frequently assess experiences and behaviors in the context of daily life. Using such techniques, one can capture momentary experiences and contextual features, including mood, thoughts, symptoms, activity, location, stress, and social events (e.g., Myin-Germeys et al., 2009; Myin-Germeys et al., 2018). The method is much less susceptible to recall bias compared to retrospective questionnaires, because experiences are assessed in the moment (Solhan et al., 2009). ESM also improves ecological validity, because the method is used in the "real world" and not in a laboratory environment (Myin-Germeys et al., 2009; Trull & Ebner-Priemer, 2009).

Some of the earliest applications of the method were focused primarily on obtaining accurate information on the time spent on various daily activities in healthy individuals in combination with assessments of the quality of these activities and the individuals' emotional states surrounding them (e.g., Brandstätter, 1983; Csikszentmihalyi et al., 1977; Larson & Csikszentmihalyi, 1978; Pawlik & Buse, 1982). However, the method was soon more widely adopted to assess experiences and contextual information in clinical (e.g., Barge-Schaapveld et al., 1995; Delespaul, 1995) and other vulnerable populations (e.g., Burton et al., 2007; Litt et al., 1998; Van Eck et al., 1996) and has more recently also been considered as part of interventional strategies (e.g., Myin-Germeys et al.,

✉ Aki Rintala
aki.rintala@kuleuven.be

1 Department of Neurosciences, Center for Contextual Psychiatry, KU Leuven, Leuven, Belgium

2 Faculty of Social Services and Health Care, LAB University of Applied Sciences, Lahti, Finland

3 Universitair Psychiatrisch Centrum, UZ Leuven, Leuven, Belgium

4 Department of Psychology, Research Group on Quantitative Psychology and Individual Differences, KU Leuven, Leuven, Belgium

5 Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University, Maastricht, the Netherlands

2016; Verhagen et al., 2016). A search for "experience sampling" OR "ecological momentary" in the Web of Science Core Collection shows an exponential increase in articles referring to such assessment techniques especially over the last 10 years.

Given the increase in the popularity of the method, it is important to recognize that the method is not without its problems. During an ESM study, participants receive multiple prompts or diary signals (hereafter also referred to as "beeps") per day to fill out a questionnaire, which makes ESM a demanding assessment tool. Hence, researchers planning the use of such data collection methods should consider the purely 'objective' (time) burden placed on the study participants. Although the questionnaire is typically kept relatively short, participants may be asked to fill in the questionnaire 50 or more times over the course of a study (Vachon et al., 2019; Van Roekel et al., 2019; Wen et al., 2017). Even if the completion time for each individual assessment is on the order of a few minutes, the total amount of time for all assessments combined can easily accumulate to several hours.

Moreover, on a more subjective level, the repeated prompts to fill in the questionnaire may also be experienced as disruptive, disturbing, irritating, or inconvenient by the study participants (Bolger et al., 2003), who may become frustrated when they are frequently asked to stop their daily life activities to answer the questionnaire (Hufford, 2007). This may lead participants to engage in various strategies to reduce such burden, for example by ignoring beeps (i.e., non-compliance), quickly responding to the questions without reflection (i.e., careless responding), or dropping out of the study altogether (i.e., attrition), which in turn negatively impacts data quantity and quality. It is therefore important to understand the extent to which participants in an ESM study experience such reactions to the beeps, as to better gauge whether ESM is an acceptable tool to assess momentary experiences in daily life.

Clearly, researchers are not observing massive attrition during their ESM studies, as otherwise this would have led many to abandon the use of this data collection method altogether, although 'failed' studies may simply not be published. A recent meta-analysis (Vachon et al., 2019), however, also found high retention rates across 79 ESM studies and no evidence for publication bias in the reported rates. Still, participants may find the experience of participating in such research as burdensome.

Some early evidence in this regard was provided by Stone et al. (2003), who randomized 91 chronic pain patients to four different conditions: Participants received either 3, 6, or 12 prompts per day for two weeks to fill in an 18-item questionnaire or were assigned to a control condition without momentary assessments. After the two weeks, participants were asked to fill in a questionnaire assessing their experience in the study, which included questions about the level of perceived burden and interference with daily activities due to study participation. Especially the group receiving 12 beeps per day reported significantly higher levels of burden and interference (with averages of 1.33 and 1.25 on a 0–3 scale, respectively) compared to groups receiving a lower number or no repeated assessments (with ranges of 0.43 to 0.91 and 0.50 to 0.64 for burden and interference, respectively). The high-frequency assessment group also reported a higher level of 'annoyance' with the repeated assessments (a mean of 1.58) compared to the 6- and 3-beep conditions (with means of 0.45 and 0.23, respectively).

More recently, Edwards et al. (2016) conducted a study to examine differences between a group of healthy controls (n = 58) and a group of individuals diagnosed with schizophrenia (n = 53) with respect to the experience and acceptability of participating in ESM research. In this study, participants were asked to fill in a questionnaire 7 times per day over the course of 6 days. When asked about disruptions to typical routines or activities afterwards, both groups reported relatively low levels (means of 2.48 and 2.23, respectively, rated on a 1–7 scale), but the degree of experienced disruption correlated with compliance in the schizophrenia group. A similar feasibility study, conducted by Jones et al. (2016), examined various aspects of participation acceptability in a group of 100 African-American women who were asked to fill in a questionnaire 5 times per day for 7 days. Ratings of the inconvenience of completing the questionnaire and interference with daily activities were also relatively low (means of 2.33 and 2.42, respectively, rated on a 1–5 scale) but again higher inconvenience was associated with lower compliance.

A few additional studies have also assessed aspects related to burden, disruption, or inconvenience as part of the questionnaire administered at every beep. Van Genugten et al. (2020) investigated such momentary burden (using the item "How disturbing is filling out the questionnaire right now?" rated on a 1–7 scale) in an ESM protocol of 14 days with 5 beeps per day in a sample of 384 participants with an affective disorder (n = 100), remitted persons (n = 190), and healthy volunteers (n = 94). Participants with an affective disorder reported on average significantly higher levels of burden (with a median of 1.98) than those from the remitted (median of 1.64) and the healthy control groups (median of 1.28). Similarly, in two studies described by van Roekel et al. (2019) with early adolescents (n = 303 and n = 241, respectively), the first running for 6 days with 9 beeps per day and the second study for 7 days with 8 beeps per day, the questionnaire that participants were asked to fill out at each beep included the item "I was inconvenienced by this beep" rated on a 1–7 scale. Participants reported moderate (average 4.1) levels of inconvenience in the first and relatively low (average of 2.2) levels in the second study. The authors noted that this difference may be due to the less intrusive notification sound used in the second study. Finally, Eisele et al. (2020) investigated momentary burden (based on the average of the items "This beep

disturbed me" and "Filling in the questionnaire took effort", each rated on a 1–7 scale) in a group of students ($n = 164$) that were randomized to receive either a 30- or 60-item questionnaire 3, 6, or 9 times per day for 14 days. The average level of momentary burden was relatively low in all groups, ranging from 2.2 to 3.0, but significantly higher for those receiving the longer instead of the shorter questionnaire (2.9 vs 2.3, respectively).

While these findings suggest that the level of burden and inconvenience reported by study participants has been relatively low in previous ESM studies, they indicate that there may be groups for whom or conditions under which participation in ESM research may be experienced as more burdensome or disruptive. Hence, aside from the overall level of burden or disturbance, an equally pertinent question is whether the degree of disturbance is associated with certain person-level characteristics and/or the particular circumstances surrounding a beep. In fact, some evidence in this regard was provided by van Genugten et al. (2020) who found that the level of disturbance reported by the participants was associated with their emotional state (i.e., positive and negative affect) at the time of the beep. Furthermore, additional analyses of the two studies by van Roekel et al. (2019) indicated that the beeps were experienced as more inconvenient when participants were in a public place as opposed to when at home or in school (in study 1), for beeps that occurred on the weekends (in study 2), and when participants were alone (in both studies). Moreover, when not alone, the level of inconvenience also depended to some extent on the type of company. In particular, beeps were experienced as more inconvenient when participants were in the presence of less familiar people (e.g., acquaintances) (Van Roekel et al., 2019).

The purpose of the present study was to further examine to what extent participants in ESM research report feelings of disturbance by the repeated prompts to fill out the questionnaire using a pooled sample of 1319 adults from 7 different ESM studies, including both healthy participants and individuals with different mental health conditions. Furthermore, we systematically examined associations between the level of reported disturbance with various person-, state-, and study-level characteristics and expand on previous research by analyzing a much broader set of associations. Following prior studies, we examined associations with the location, the social context, and the emotional state at the time of the beep, but also investigated various stress-related characteristics, as these have received considerable attention in prior research in themselves (e.g., Collip et al., 2011; Lataster et al., 2010; Myin-Germeys et al., 2001; Myin-Germeys et al., 2003; Myin-Germeys & Van Os, 2007). In particular, it is plausible that participants may experience the beeps as more disturbing if they are in a certain emotional state (e.g., moments characterized by higher levels of negative affect or stress). Furthermore, we also considered possible associations with the physical state (e.g., feeling tired, hungry, or unwell) of the participants at the time of the beep and the use of certain substances, as these have been found to be relevant predictors of compliance (Rintala et al., 2020) and hence may also be predictive of the level of disturbance. Like van Genugten et al. (2020) and Eisele et al. (2020), we also examined possible changes in the level of disturbance over time, but also tested for changes as a function of the time within each day. We also attempted to replicate the finding by van Roekel et al. (2019) that beeps may be experienced as more disturbing during the weekends. Finally, some of the studies included in the analysis asked participants to provide a saliva sample at each beep (to measure cortisol levels). Doing so could potentially be experienced by participants as more burdensome than just responding to the questionnaire, leading to higher levels of reported disturbance in such studies.

Methods

Participants

Analyses were conducted using a pooled dataset of 7 studies comprising a total of 1354 participants (see Supplementary file 1 for more details on the included studies). From the 1354 participants, sufficient data for inclusion in the analysis were available from 1319 (97.4%) participants. Twenty-six participants were excluded due to missing data in one or multiple variables at each beep.[1] Four participants were excluded due to missing data in age, three in mental health status, and one in gender. From the included 1319 participants, beeps where responses to all variables of interest (see below) were available were included in the final analysis (40,382 observations). The final sample comprised 973 (73.8%) female and 346 (26.2%) male subjects with a mean (SD, range) age of 32.1 (11.0, 16–65) years. Among the participants, 797 (60.4%) were classified as healthy subjects, 190 (14.4%) were suffering from some form of psychosis, 129 (9.8%) from depression, 176 (13.3%) had a familial risk for psychosis (i.e., having a first-degree relative with a psychotic disorder), and 27 (2.1%) had a psychometric risk for psychosis (i.e., persons scoring high on a subclinical psychosis scale). The local ethics committees approved all studies included in the pooled dataset and written informed consent was obtained from all participants.

[1] These participants missed different items at each responded beep for unknown reasons. For example, one subject filled in affect-related items on some occasions, but did not fill in stress- or context-related items. In other words, these participants consistently lacked either one or several beep-level variables of interest in the dataset.

ESM Protocol

All 7 studies in the pooled dataset used an identical ESM protocol for either 4, 5, or 6 consecutive days using a paper-and-pencil diary and a digital wristwatch. Participants received 10 randomized beeps per day in 90-min intervals between 7.30 a.m. and 10.30 p.m.. More details on the questionnaires used in the studies is published elsewhere (Rintala et al., 2019).

The set of questions included in the ESM questionnaires differed slightly across studies and ranged from 42 to 52 questions, but all questionnaires included items assessing emotions, physical state, activity, location, and appraisals of the current situation. Most items in the questionnaires were scored on a 7-point Likert scale (e.g., "I feel cheerful" with 1 = "not at all" to 7 = "very much"). A few questions were open-ended (e.g., "What am I doing?") or were scored on a bipolar scale (e.g., the event-related question "This event was" -3 = "very unpleasant" to +3 = "pleasant") or a binary scale (e.g., "I am alone" with response options "Yes" or "No"). An example questionnaire is provided in Supplementary file 2.

Measures

Disturbance by the Beep

Disturbance by the beep was assessed with the item "*This beep disturbed me",* scored on a 7-point Likert scale ranging from 1 ("not at all") to 7 ("very much"). Participants received this item at every beep as the last item in the questionnaire. To report the frequency of disturbance, we divided the variable into four levels (1 = no disturbance, 2–3 = low, 4–5 = moderate, and 6–7 = high disturbance).

Predictors of Interest

Personal Characteristics Three variables were extracted from the dataset: age, gender, and study population. Age was treated as a continuous variable (coded (age – 20) / 10 to avoid an overly small coefficient), and gender was coded as "0" for males and "1" for females. Study population was examined as a five-level factor according to the classification described earlier, with healthy participants used as the reference category.

A potential concern with the present data is that study participants may be more likely to ignore or miss beeps under circumstances where disturbance levels are especially high. As a result, at least the estimate of the overall level of disturbance may be biased downward. Since we cannot examine the association between disturbance levels and missingness at the beep level (since the level of disturbance is missing/unknown for beeps not responded to), we computed both the person- and day-level compliance (as proportions) for each participant and each day and included them in the model as proxies.

Emotional Characteristics Participants rated their positive (PA) and negative (NA) affect on a 7-point Likert scale, ranging from 1 to 7. PA was computed based on the mean of the three items "*cheerful*", "*satisfied*", and "*relaxed*", while NA was based on the mean of the six items "*uncertain*", "*lonely*", "*anxious*", "*guilty*", "*down*", and "*irritated*". The items for these composites were chosen because they were assessed in all of the included studies (the set of emotional characteristic items included in the various studies varied slightly) and because previous ESM studies have used the same or similar items to assess PA and NA (Collip et al., 2013; Myin-Germeys et al., 2001). Reliabilities (i.e., internal consistency) of these composites at the person and beep levels were 0.91 and 0.76 for PA and 0.92 and 0.69 for NA, respectively.

Stress-Related Characteristics Activity stress was defined according to previous ESM studies (Lataster et al., 2010; Myin-Germeys et al., 2001), which was assessed through the mean of the three items "*I am skilled to do this activity*" (reverse scored), "*I prefer doing something else*", and "*This is a challenge*" that were each rated on a 7-point Likert scale.

For event pleasantness, participants were asked to recall the most important event that happened between the current and the previous beep. This event was then rated on a 7-point bipolar scale (−3 = very unpleasant to +3 = very pleasant). The item was reverse coded so that higher scores reflect higher unpleasantness. Previous ESM studies have used this item to indicate "*event-related stress*" (Collip et al., 2011; Lataster et al., 2013) and we will adopt this terminology as well.

Contextual Characteristics We defined location and the social situation as contextual characteristics. Location was dummy coded, where "0" indicated being at home and "1" being somewhere else. Social situation was coded as a 7-level factor using six dummy variables, where being alone was treated as the reference category. The six dummy variables were coded as "1" for "*family and friends*", "2" for "*acquaintances*", "3" for "*strangers*", "4" for a combination of "*family and friends"* with "*acquaintances"*, "5" for a combination of "*family and friends"* with "*strangers*", and "6" for a combination of "*acquaintances*" and "*strangers*".

Physical State Characteristics Four items that reflect a participant's physical state were included, namely "*I am active*", "*I am unwell*", "*I am tired*", and "*I am hungry*" that were all rated on a 7-point Likert scale.

Substance Use Characteristics Participants were asked to fill in items related to the use of substances since the previous beep.

Dummy variables were coded for tobacco, alcohol, and medication use ("0" No, "1" Yes).

Time Characteristics Time characteristics consisted of three different variables: Chronological study day (i.e., 1 through 4, 5, or 6), calendar day, and time within a day (i.e., the beep number within a given day from 1 to 10). For the analyses, chronological study day was examined as a six-level factor using the first study day as the reference category. For calendar day, we distinguished between weekdays (coded as "0") and weekends (coded as "1"). Time of the day was coded as a 10-level factor taking the first beep of the day (i.e., between 7.30 a.m. and 9 a.m.) as the reference category.

Study Characteristics We examined one study characteristic, namely whether studies used saliva sampling at every assessment or not. Three out of the 7 studies in our pooled dataset measured cortisol levels from saliva samples at each beep. A dummy variable was coded as "0" for studies that did not use saliva sampling and "1" for studies that used saliva sampling.

Analyses

Analyses were conducted using multilevel (mixed-effects) regression models with dichotomous, categorical, and continuous beep-, day-, and study-level predictor variables and the reported level of disturbance at each beep as the outcome variable. First, we fitted an intercept-only model to investigate the overall level of disturbance by the beeps across studies. Next, we fitted a multivariable model to examine the association between the predictor variables of interest and the level of disturbance. Both models included random effects for studies, subjects within studies, study days within subjects, and beep number within subjects, with the last two entered as crossed random effects. In the multivariable model, the beep-level variables were included as between-subject (BS) variables (by entering the subject-level means) and as within-subject (WS) variables (by entering the deviations from the subject-level means) simultaneously. This allowed us to investigate differences in disturbance between individuals that for example report on average higher versus lower levels of negative mood overall (BS) and changes in disturbance when a person reports different levels of negative mood at particular moments (WS). For each of the substance use characteristics, we also included an additional subject-level dummy variable in the model to distinguish those who had never consumed a particular substance at any of the beeps versus those who had. For example, for alcohol, this allowed us to examine differences between total abstainers versus those who had consumed alcohol at least once during the study period (BS_1), in addition to examining whether disturbance levels differed between heavier versus less heavy consumers (BS_2) and when comparing moments when a non-abstainer had consumed alcohol versus not (WS).

Based on the intercept-only model, we estimate the average level of disturbance, the corresponding 95% confidence interval (95% CI), and intraclass correlation coefficients (ICCs) for the variance components (i.e., the estimated proportions indicating how much each variance component contributes to the total variance in the model). For the multivariable model, we report the estimated intercept and slope for each predictor with corresponding Wald-type z-tests. Factors as a whole were tested with Wald-type chi-square tests. Analyses were conducted using R 4.0.3 (R Development Core Team, 2016) with packages lme4 (Bates et al., 2015) and car (Fox & Weisberg, 2011).

Results

An overview of the variables of interest is presented in Table 1. The results from the multivariable model are presented in Table 2.

Overall Disturbance by the Beep

Participants did not perceive any level of disturbance in 36.5% of the beeps. Low disturbance was reported in 25.6% of the beeps, moderate in 26.4%, and the highest level of disturbance was reported in 11.5% of the beeps. Based on the intercept-only model, the estimated average level of disturbance was 2.84 (95% CI [2.60, 3.07]). Out of the total variance, 55% was due to moment-to-moment fluctuations in the disturbance levels, 34% was due to subject-level differences, and the remaining 11% was due to the other components (subject and day number: 7%; subject and beep number: 2%; study: 2%). Hence, study-level differences were relatively small, giving some assurance that the combination of the studies into a single dataset is defensible.

Predictors of Disturbance

Personal Characteristics Younger participants reported higher levels of disturbance ($b = -0.298$, $Z = -8.548$, $p < .0001$). On the other hand, gender ($b = 0.065$, $Z = 0.765$, $p = .45$), study population ($\chi^2(4) = 4.8755$, $p = .30$), person-level compliance ($b = -0.123$, $Z = -0.752$, $p = .45$), or day-level compliance ($b = -0.048$, $Z = -0.975$, $p = .33$) were not significantly associated with the level of disturbance.

Emotional Characteristics Higher levels of PA were associated with lower disturbance at both the between- and within-subject levels (BS: $b = -0.161$, $Z = -3.307$, $p = .001$; WS: $b = -0.134$, $Z = -13.270$, $p < .0001$). In contrast, higher levels

Table 1 Descriptive information of the participants and the self-reported momentary experiences and situations ($N = 1319$)

	Observations	Percent or mean (SD)
Participants		
Gender		
Female	973	73.8%
Male	346	26.2%
Age		32.10 (11.03)
Clinical status		
Healthy participants	797	60.4%
Psychosis	190	14.4%
Depression	129	9.8%
Familial risk for psychosis	176	13.3%
Psychometric risk for psychosis	27	2.1%
Self-reported momentary experiences and situations		
Disturbance of the ESM beep (range of 1–7)		
Not at all (response of 1)	14,732	36.5%
Low disturbance (responses of 2 & 3)	10,332	25.6%
Moderate disturbance (responses of 4 & 5)	10,691	26.4%
High disturbance (responses of 6 & 7)	4627	11.5%
Positive affect	40,382	4.71 (1.29)
Negative affect	40,382	1.52 (0.83)
Activity stress	40,382	2.65 (1.11)
Event-related stress	40,382	2.72 (1.63)
Location		
Being at home	23,121	57.3%
Being outside home	17,261	42.7%
Social situation		
Alone	14,415	35.7%
With family and friends	17,744	43.9%
With acquaintances	5205	12.9%
With strangers	1341	3.3%
With family, friends, and acquaintances	687	1.7%
With family, friends, and strangers	430	1.1%
With acquaintances and strangers	560	1.4%
Being active	40,382	3.51 (1.96)
Being physically tired	40,382	2.81 (1.77)
Being physically unwell	40,382	1.76 (1.34)
Being hungry	40,382	2.06 (1.64)
Tobacco use		
No	31,903	79.0%
Yes	8479	21.0%
Alcohol use		
No	38,193	94.6%
Yes	2189	5.4%
Medication use		
No	38,204	94.6%
Yes	2178	5.4%

SD, Standard deviation

of NA were associated with higher disturbance only at the within-subject level ($b = 0.060$, $Z = 3.600$, $p = .0003$).

Stress-Related Characteristics Higher levels of activity stress were associated with higher disturbance at both the between-

Table 2 Results of the multivariable model examining predictors of the perceived level of disturbance by the beeps (N = 1319 and 40,382 observations)

		b	Z	*p*
	Model intercept	1.963	4.664	
Personal characteristics				
Age		−0.298	−8.548	**< .0001**
Gender				
	Male[§]			
	Female	0.065	0.765	.445
Study population				
	Healthy participants[§]			
	Psychosis	−0.188	−1.568	.118
	Depression	0.126	0.569	.602
	Familial risk for psychosis	0.046	0.431	.667
	Psychometric risk for psychosis	−0.156	−0.591	.556
	($\chi^2(4)=4.876$; $p=.300$)			
Person-level compliance		−0.123	−0.752	.452
Day-level compliance		−0.048	−0.975	.329
Emotional characteristics				
Positive affect				
Between subjects		−0.161	−3.307	**.001**
Within subjects		−0.134	−13.270	**< .0001**
Negative affect				
Between subjects		0.139	1.894	.058
Within subjects		0.060	3.600	**.0003**
Stress-related characteristics				
Activity stress				
Between subjects		0.235	4.057	**< .0001**
Within subjects		0.032	3.805	**.0001**
Event-related stress				
Between subjects		0.090	1.901	.057
Within subjects		−0.008	−1.615	.106
Contextual characteristics				
Location				
	Being at home[§]			
Between subjects	Being outside home	−0.605	−4.324	**< .0001**
Within subjects	Being outside home	0.249	12.477	**< .0001**
Social situation				
	Alone[§]			
	Family and friends	0.172	9.001	**< .0001**
	Acquaintances	0.218	7.399	**< .0001**
	Strangers	0.363	7.842	**< .0001**
	Family, friends, and acquaintances	0.269	4.467	**< .0001**
	Family, friends, and strangers	0.326	4.300	**< .0001**
	Acquaintances and strangers	0.607	8.866	**< .0001**
	($\chi^2(6)=185.001$; $p<.0001$)			
Physical state characteristics				
Being active				
Between subjects		0.125	3.818	**.0001**
Within subjects		0.080	16.884	**< .0001**

Table 2 (continued)

		b	Z	*p*
	Model intercept	1.963	4.664	
Feeling tired				
Between subjects		0.142	4.161	**< .0001**
Within subjects		0.058	9.565	**< .0001**
Feeling unwell				
Between subjects		−0.024	−0.518	.605
Within subjects		−0.019	−2.260	**.024**
Feeling hungry				
Between subjects		0.058	1.268	.205
Within subjects		0.032	6.277	**< .0001**
Substance use characteristics				
Tobacco use				
	Never/No§			
At least once	Yes	0.153	1.343	.180
Between subjects	Yes	−0.410	−2.659	**.008**
Within subjects	Yes	−0.172	−4.534	**< .0001**
Alcohol use				
Never/No§				
At least once	Yes	0.140	1.753	.080
Between subjects	Yes	0.071	0.152	.879
Within subjects	Yes	0.020	0.575	.565
Medication use				
	Never/No§			
At least once	Yes	0.079	0.923	.356
Between subjects	Yes	0.130	0.288	.773
Within subjects	Yes	−0.144	−4.047	**< .0001**
Time characteristics				
Chronological study days				
	1§			
	2	0.060	1.980	**.048**
	3	0.130	4.194	**< .0001**
	4	0.169	5.379	**< .0001**
	5	0.192	6.012	**< .0001**
	6	0.190	5.034	**< .0001**
	($\chi^2(5)=56.406$; $p<.0001$)			
Beep number (time of the day)				
	1 (7.30 a.m. – 9 a.m.)§			
	2 (9 a.m. – 10.30 a.m.)	−0.216	−5.933	**< .0001**
	3 (10.30 a.m. – 12 p.m.)	−0.258	−7.110	**< .0001**
	4 (12 p.m. – 1.30 p.m.)	−0.145	−4.044	**< .0001**
	5 (1.30 p.m. – 3 p.m.)	−0.182	−4.988	**< .0001**
	6 (3 p.m. – 4.30 p.m.)	−0.119	−3.234	**.001**
	7 (4.30 p.m. – 6 p.m.)	−0.055	−1.502	.133
	8 (6 p.m. – 7.30 p.m.)	−0.051	−1.397	.163
	9 (7.30 p.m. – 9 p.m.)	−0.061	−1.656	.098
	10 (9 p.m. – 10.30 p.m.)	−0.103	−2.715	**.007**
	($\chi^2(9)=94.354$; $p<.0001$)			

Table 2 (continued)

		b	Z	*p*
	Model intercept	1.963	4.664	
Day of week				
	Weekdays[§]			
	Weekend	0.068	3.133	**.002**
Study characteristics				
Saliva sampling				
	No[§]			
	Yes	0.234	1.537	.221

Significant results are displayed in bold; *b* = estimate value of coefficient; Z = Z-value; *p* = *p* value; [§] = Reference category; χ^2 = Wald-type chi-square tests to test the categorical variable as a whole

and within-subject levels (BS: $b = 0.235$, $Z = 4.057$, $p < .0001$; WS: $b = 0.032$, $Z = 3.805$, $p = .0001$). On the other hand, event-related stress was not significantly associated with disturbance (BS: $b = 0.090$, $Z = 1.901$, $p = .06$; WS: $b = -0.008$, $Z = -1.615$, $p = .11$).

Contextual Characteristics Participants who were away from their home more frequently reported lower disturbance overall (BS: $b = -0.605$, $Z = -4.324$, $p < .0001$). However, the opposite finding was observed at the within-subject level (WS: $b = 0.249$, $Z = 12.477$, $p < .0001$). Disturbance increased when participants were in the presence of any type of social company ($\chi^2(6) = 185.001$, $p < .0001$). This effect was most pronounced in the presence of acquaintances and strangers ($b = 0.607$, $Z = 8.866$, $p < .0001$) or strangers alone ($b = 0.363$, $Z = 7.842$, $p < .0001$).

Physical State Characteristics Higher levels of activity were associated with higher disturbance at both the between- and within-subject levels (BS: $b = 0.125$, $Z = 3.818$, $p = .0001$; WS: $b = 0.080$, $Z = 16.884$, $p < .0001$). Participants reporting higher levels of tiredness overall and at particular moment (BS: $b = 0.142$, $Z = 4.161$, $p < .0001$; WS: $b = 0.058$, $Z = 9.565$, $p < .0001$) also perceived the beeps as more disturbing. Feeling unwell was associated with lower disturbance (WS: $b = -0.019$, $Z = -2.260$, $p = .024$) and feeling hungry was associated with higher disturbance (WS: $b = 0.032$, $Z = 6.277$, $p < .0001$) at the within-subject level.

Substance Use Characteristics While there was no significant difference between those who consumed tobacco at least once versus those who never did (BS_1: $b = 0.153$, $Z = 1.343$, $p = .18$), participants who used more tobacco overall or who had used tobacco at particular moments reported lower disturbance (BS_2: $b = -0.410$, $Z = -2.659$, $p = .008$; WS: $b = -0.172$, $Z = -4.534$, $p < .0001$). Medication use was also negatively associated with disturbance at the within-subject level (WS: $b = -0.144$, $Z = -4.047$, $p < .0001$). In contrast, any alcohol use (BS_1: $b = 0.140$, $Z = 1.753$, $p = .08$), heavier use overall (BS_2: $b = 0.071$, $Z = 0.152$, $p = .88$), or use at particular moments (WS: $b = 0.020$, $Z = 0.575$, $p = .57$) was not significantly associated with disturbance levels.

Time and Study Characteristics Disturbance increased across chronological study days ($\chi^2(5) = 56.406$, $p < .0001$; Fig. 1a), reaching the highest level on the fifth study day ($b = 0.192$, $Z = 6.012$, $p < .0001$). On the sixth study day, disturbance seemed to stabilize ($b = 0.190$, $Z = 5.034$, $p < .0001$). Disturbance levels also varied significantly within a day ($\chi^2(9) = 94.354$, $p < .0001$; Fig. 1b), with the highest level reported at the 1st beep of the day. The level of disturbance was the lowest on the 3rd beep ($b = -0.258$, $Z = -7.110$, $p < .0001$). However, disturbance increased again towards the evening and did not differ from the first beep at the 7th beep ($b = -0.055$, $Z = -1.502$, $p = .13$), 8th beep ($b = -0.051$, $Z = -1.397$, $p = .16$), and 9th beep ($b = -0.061$, $Z = -1.656$, $p = .10$). Participants also perceived higher levels of disturbance during the weekends compared to weekdays ($b = 0.068$, $Z = 3.133$, $p = .002$). On the other hand, ESM protocols using saliva sampling at each beep did not lead to significantly different levels of disturbance compared to studies without saliva sampling ($b = 0.234$, $Z = 1.537$, $p = .22$).

Discussion

Overall, participants reported relatively low levels of disturbance by the beeps (mean of 2.84 on a 1–7 scale). For most beeps, participants perceived either no disturbance at all or what could be considered low to moderate levels of disturbance. These findings suggest that a high frequency ESM protocol with 10 beeps per day can be considered an alternative protocol to use in ESM research. However, aside from the overall level of disturbance, an equally important question is whether the degree of disturbance is associated with certain person-level characteristics and/or the particular circumstances surrounding a beep, which may reveal how participants perceive the beeps for example at different times during the day or when experiencing different states or daily life situations.

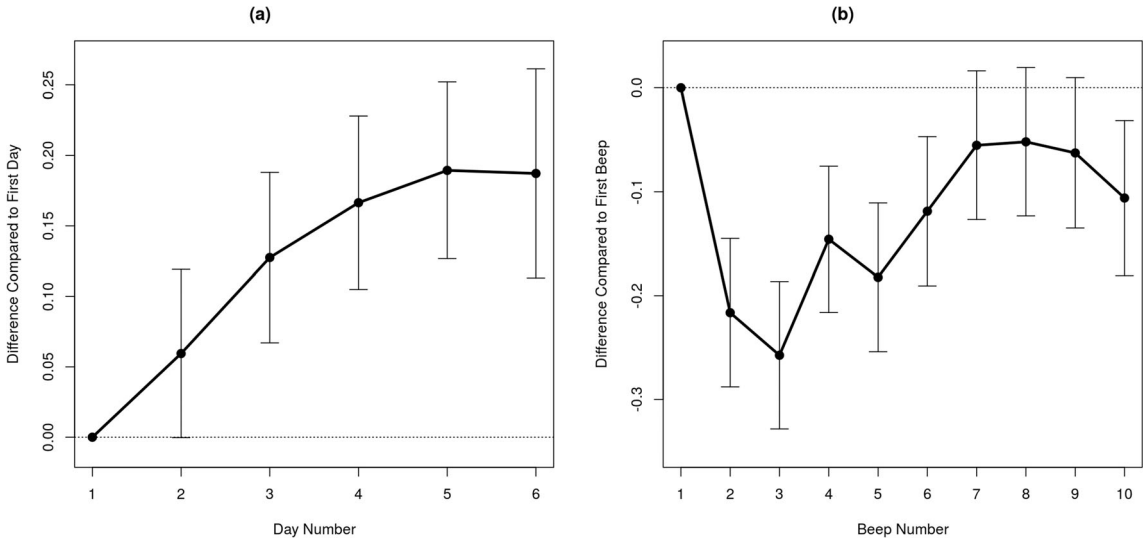

Fig. 1 Difference in the level of reported disturbance when compared against (**a**) the first study day and (**b**) the first beep of the day based on the multivariable model (with 95% confidence intervals)

Age, Gender, Study Population, and Compliance

Younger participants in our sample tended to report higher levels of disturbance. For example, if we compare 50 versus 20 year olds (which corresponds approximately to the 90th versus 10th percentile of the participants' ages), the findings imply that the disturbance level reported by the younger participants was on average 0.90 points higher. This is a rather sizable difference on a 1–7 scale and represents a new finding in ESM research, as previous ESM studies have not investigated age as a predictor of momentary burden, inconvenience, or disturbance. Given that the model controlled for various other predictors, such as levels of activity and social interactions, this finding cannot be explained for example by younger participants being more active or outgoing (assuming that participants would experience the beeps as more disturbing under such circumstances). However, age may be related to other aspects that we could not control for in our analyses (e.g., hobbies or work circumstances) which may explain this association, but this will require further investigation in the future.

Neither gender nor clinical population was associated with the level of reported disturbance. The lack of a difference between the clinical status groups is in line with the findings by Edwards et al. (2016) who also did not detect statistically significant differences between people with schizophrenia and healthy controls in the level of disruption to typical routines or activities experienced by the study participants. However, van Genugten et al. (2020) observed that persons with a current affective disorder (i.e., persons who met the criteria for a depressive or anxiety disorder in the past six months) reported higher momentary burden compared to remitted and healthy participants. While we also found that participants suffering from depression reported on average the highest levels of disturbance (0.15 points higher compared to the healthy controls), this difference was not statistically significant. Although this is purely speculative, a possible explanation might be the relatively high prevalence of anxiety disorders in the sample by van Genugten et al. (2020), assuming that such disorders are associated with a heightened sense of disruption. Given these somewhat conflicting findings, further studies are recommended to examine whether differences in momentary burden and disturbance exist in different clinical populations and when compared to healthy participants.

Concerns about the fact that the analyses can only make use of those beeps to which participants responded are alleviated (although not eliminated) by not finding significant associations between the person- and day-level measures of compliance and the reported disturbance levels. Although associations may differ across levels, if the chances of missingness are indeed higher at moments where the beep is experienced as highly disturbing, we might expect to see lower disturbance levels in less compliant participants or on days where participants are less compliant. The lack of evidence for these associations suggests that participants may not be engaging in such a strategy to reduce burden.

PA and NA

Participants with higher PA overall tended to report lower levels of disturbance. Also, how disturbing the beeps were experienced in particular moments was negatively associated with the level of PA and positively associated with the level of NA at those moments. Strictly speaking (i.e., with $p = .057$), we did not find that participants with higher NA overall tended to report higher disturbance levels, but the unique effects of PA and NA are difficult to differentiate due to the relatively strong correlation between these constructs (the subject-level means of PA and NA correlated −0.54 in our

sample). Broadly, these findings are therefore in line with those by van Genugten et al. (2020) who also found that the level of disturbance reported by participants was associated with their emotional state at the time of the beep.

Our analyses need to be interpreted with caution though, as we only examined associations at concurrent moments. Therefore, we cannot conclude any causality or directionality based on our findings. Participants may feel more disturbed by the beep when they are already in an emotional state characterized by low PA and high NA or the ESM signal may at times actually trigger or at least contribute to the presence of such emotional states. The former may seem more plausible, but we cannot rule out that the beep signal itself may affect the participants' emotional states or that other factors not accounted for in the model are the underlying reason for this association.

Activity and Event-Related Stress

Participants who reported higher levels of activity stress overall and in particular moments tended to report higher levels of disturbance. Not surprisingly, when participants are asked to fill in the ESM questionnaire while engaged in some stressful or challenging activity, this appears to be experienced as particularly disruptive. However, similar to the finding of the association between the emotional states (i.e., PA and NA) and disturbance, we cannot rule out the possibility that the interruption itself contributes to the current activity as being experienced as more stressful. On the other hand, for event-related stress, we did not find evidence that this variable is associated with the level of disturbance. The absence of a significant association may at least be partly explained by the fact that participants were asked to reflect on the most important event that happened between the current and the previous beep. Due to the time separation, any effect that a particularly unpleasant event might have had on the disturbance level may have dissipated in the meantime.

Location and Social Company

Participants who spent more time outside their home tended to report lower levels of disturbance on average, but in particular moments being outside the home was associated with higher disturbance levels. Apparently, some person-level characteristic of more 'outgoing' individuals is associated with a higher acceptability of the ESM protocol. On the other hand, the results also indicated that the prompts are still experienced as more disturbing when they occur outside the home, which is at least partly supported by the findings of the study by van Roekel et al. (2019) with adolescents who experienced higher inconvenience by the beeps when in public surroundings compared to at home, although the lowest levels of inconvenience were reported when at school. The opposite directionality of the association at the between- and within-subject levels in our study highlights the importance of properly disentangling the effects at these two levels.

Participants who were in company experienced the prompts as more disturbing when compared to situations where they were alone, a finding opposite to that of van Roekel et al. (2019). However, when looking at the level of disturbance in different social situations, participants in the company of less familiar persons (i.e., strangers and/or acquaintances) reported higher disturbance levels compared to social situations with more familiar persons (i.e., family and/or friends), which is again in line with the results reported by van Roekel et al. (2019). Especially the presence of strangers appears to be a trigger for higher disturbance levels in our study, leading to an increase of more than 0.3 points on average when compared to being alone. While one might suspect that this finding is related to the specific data collection method used in the included studies (i.e., filling in a paper-and-pencil diary might be experienced as somewhat awkward in such moments), the use of smartphones by van Roekel et al. (2019) led to similar findings. This effect might therefore be less driven by the presence of strangers per se, but more so by the circumstances characterized by their presence (e.g., while running errands, taking public transportation) where filling out the questionnaire might be experienced as particularly inconvenient.

Physical States

Participants who reported higher levels of activity and tiredness on average and also in a given moment tended to report higher levels of disturbance. We also found such an association for feeling hungry at particular moments. Fluctuations in these physical states are a natural part of daily life and not surprisingly participants appear to perceive the prompts to fill out a questionnaire in moments where these states are elevated as more bothersome.

Counterintuitively, we found that participants feeling unwell at particular moments experienced the prompts as less disturbing, although the association was weak and just barely statistically significant. While the dataset did not encompass further details on the circumstances surrounding such moments, it is plausible that persons may then choose to be in an environment where filling out the questionnaire is not experienced as disturbing (e.g., while resting).

Substance Use

Although we found no significant differences between those who consume certain substances (i.e., tobacco and alcohol) or who use medication versus those who do not, heavier users of tobacco and participants who reported the use of tobacco or medication since the last beep tended to report lower levels of

disturbance. For tobacco users, non-consumption is presumably a trigger for being in a somewhat heightened state of agitation and hence an increased perception of disturbance by an ESM prompt. Although we have no additional information on the types of medications used by the participants, almost 25% of the participants either had a diagnosis of some form of psychosis or depression, the majority of which (i.e., 64%) indicated the use of medication during the study period compared to only 28% of the healthy participants. At least for these participants, medication use may be an important determinant for their mental state at particular moments, potentially explaining why medication use is associated with lower disturbance levels at the subsequent beep.

Study Days, Time within a Day, Weekdays, and Saliva Sampling

We found that the level of disturbance increased across study days by 0.06 to almost 0.20 points on average when compared to the first study day, although this trend seemed to flatten out around the 5th day. Previous studies that modeled changes in disturbance over time linearly also found positive coefficients, but those did not reach statistical significance (Eisele et al., 2020; van Genugten et al., 2020). To that, we asked Eisele et al. (2020) and van Genugten et al. (2020) to reanalyze study time (i.e., study day) as a categorical variable to replicate our analyses. Also, we analyzed the data by van Roekel et al. (2019) (which the authors have shared via the Open Science Framework) to examine changes in their inconvenience measure as a function of time in the same manner. After replicating our approach, Eisele et al. (2020) again found that disturbance levels tended to be higher for latter days (for 10 out of the 13 days when compared to the first), but differences were small and not significant (G. Eisele, personal communication, November 21, 2020). However, the reanalysis by van Genugten et al. (2020) and our analysis of the data for both studies by van Roekel et al. (2019) yielded similar significant trends as in our study (C. van Genugten, personal communication, November 13, 2020). In fact, the first study by van Roekel et al. (2019) yielded a pattern that looked remarkably similar to the one we show in Fig. 1a, although the trends in the second study and that by van Genugten et al. (2020) showed more linear increases. Regardless, by treating day as a categorical variable, more insight can be gained where and how the change occurs when investigating disturbance, perceived momentary burden, or inconvenience as a function of time.

Disturbance levels were highest in the morning (between 7.30 a.m. and 9 a.m.) but then dropped significantly (by up to 0.26 points) for the subsequent two beeps covering the time intervals up to noon. Levels then started to rise again towards the afternoon and evening, almost reaching comparable levels to that of the morning for beeps between 4.30 p.m. and 9 p.m.. The highest level of disturbance at the first beep of the day may be a result of interruptions to the participants' morning routines as being experienced as particularly disturbing. Regardless of the specific reasons, we recommend that careful consideration is given to the timing of the first beep in each day. For example, some studies have adapted the scheduling of the beeps based on the participants' individual waking hours (e.g., Bastiaansen et al., 2020; Whalen et al., 2002).

Participants also reported a small but significant increase of 0.07 points on average in the level of disturbance during the weekends when compared to weekdays, which is in line with one of the studies reported by van Roekel et al. (2019). Most participants are likely to have different daily routines during the weekends compared to the weekdays, which are apparently perceived to be somewhat less compatible with participation in an ESM study.

Finally, we hypothesized that the addition of saliva sampling to each assessment may lead to the beeps as being experienced as more disturbing. However, despite the coefficient pointing in this direction, our analyses did not reveal a significant difference in terms of disturbance levels between studies with and without saliva sampling.

Strengths and Limitations, and Recommendations

A strength of this study is its large sample size of 1319 participants that could be achieved by pooling together the data from seven studies, using similar ESM protocols and questionnaires. Using this dataset, we were able to fit a single comprehensive model giving new insights into the level of disturbance and predictors thereof in a general adult population sample and individuals with psychosis, depression, and at risk for psychosis. At the same time, the studies we combined, some of which are somewhat older, used wristwatches for issuing the prompts and paper-and-pencil diaries for collecting the data, a method that is essentially replaced by the use of smartphones nowadays. This calls into question the generalizability of the present findings. However, as described above, when contrasting (where possible) some of our findings with those from other studies that did use smartphones or other electronic devices for data collection, we often found general agreement (with a few exceptions). Hence, the present study provides insights that we do not think should be summarily dismissed due to the specific data collection method used. At the same time, further studies will be needed to determine if all of the findings can be replicated when other data collection methods are used.

An important issue, which is beyond the purposes of the present study, are the consequences when participants experience heightened levels of disturbance due to the repeated prompts to fill in the questionnaire. Some of the predictors examined in the present paper have also been found to be predictive of (non)compliance (Courvoisier et al., 2012;

Messiah et al., 2011; Rintala et al., 2020; Sokolovsky et al., 2014). Furthermore, the reported disturbance level at one beep appears to be predictive of (non)compliance at the subsequent beep (Rintala et al., 2020). To what extent disturbance levels also impact data quality (i.e., reliability and validity) is a question to be addressed in future research.

Another recommendation for future studies would be to investigate whether the sound of the beep alters the degree of disturbance, given that van Roekel et al. (2019) mentioned that a less intrusive notification beep may have partly explained the differences in the level of inconvenience in their two studies (study 1: M = 4.1, study 2: M = 2.2), where the second study used a less intrusive notification sound. Unfortunately, it was not possible to examine such an association in our analysis since all studies in our pooled dataset used the same notification sound. If such an association could be confirmed, then this might lead to further optimizations (e.g., by reducing or alternating the sound of beep) to decrease disturbance levels.

To conclude, while the absence of (or more precisely, lack of evidence for) certain associations is reassuring for researchers designing ESM studies especially in clinically diverse populations, the findings also highlight that the repeated assessments in ESM studies can be experienced as more disturbing under certain circumstances (e.g., when positive affect is low, negative affect is high, when stressed due to some activity, when outside the home, in the presence of strangers, and/or when feeling tired or hungry). Since these types of circumstances are unavoidable when investigating experiences in daily life, researchers should fully acknowledge this issue in the study briefing before data collection commences, but also emphasize to participants the importance of their responses in moments where the motivation to fill in the ESM questionnaire might be low.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12144-021-01974-3.

Acknowledgements The authors wish to thank the research group of ESM MERGE for providing the data for this study.

Materials and/or Code Availability The materials and/or codes of the current study are available from the corresponding author on reasonable request.

Authors' Contribution A.R. conceived of the presented idea. A.R., M.W., & W.V. developed the conceptualization and methodology. A.R. and G.L. worked with the analytical methods, which W.V verified. A.R. was responsible for conducting the analyses and writing the manuscript with the supervision help from W.V. All promoters (M.W., W.V., and I.M-G.) supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Funding Aki Rintala and Inez Myin-Germeys was supported by the Remote Assessment of Disease and Relapse – Central Nervous System (RADAR-CNS) research programme from the Innovative Medicines Initiative 2 Joint Undertaking under a grant agreement number 115902. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA). Inez Myin-Germeys was funded by the Fonds voor Wetenschappelijk Onderzoek (FWO) Odysseus grant (GOF8416N).

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics Approval The local ethics committees approved all studies included in the pooled dataset.

Consent to Participate Written informed consent was obtained from all participants.

Conflict of Interest The authors declare that they have no conflict of interest.

References

Barge-Schaapveld, D. Q. C. M., Nicolson, N. A., van der Hoop, R. G., & DeVries, M. W. (1995). Changes in daily life experience associated with clinical improvement in depression. *Journal of Affective Disorders, 34*(2), 139–154. https://doi.org/10.1016/0165-0327(95)00012-C.

Bastiaansen, J. A., Kunkels, Y. K., Blaauw, F. J., Boker, S. M., Ceulemans, E., Chen, M., Chow, S. M., de Jonge, P., Emerencia, A. C., Epskamp, S., Fisher, A. J., Hamaker, E. L., Kuppens, P., Lutz, W., Meyer, M. J., Moulder, R., Oravecz, Z., Riese, H., Rubel, J., Ryan, O., Servaas, M. N., Sjobeck, G., Snippe, E., Trull, T. J., Tschacher, W., van der Veen, D. C., Wichers, M., Wood, P. K., Woods, W. C., Wright, A. G. C., Albers, C. J., & Bringmann, L. F. (2020). Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *Journal of Psychosomatic Research, 137*, 110211. https://doi.org/10.1016/j.jpsychores.2020.110211.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 51. https://doi.org/10.18637/jss.v067.i01.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*(1), 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030.

Brandstätter, H. (1983). Emotional responses to other persons in everyday life situations. *Journal of Personality and Social Psychology, 45*(4), 871–883. https://doi.org/10.1037/0022-3514.45.4.871.

Burton, C., Weller, D., & Sharpe, M. (2007). Are electronic diaries useful for symptoms research? A systematic review. *Journal of Psychosomatic Research, 62*(5), 553–561. https://doi.org/10.1016/j.jpsychores.2006.12.022.

Collip, D., Nicolson, N. A., Lardinois, M., Lataster, T., van Os, J., & Myin-Germeys, I. (2011). Daily cortisol, stress reactivity and psychotic experiences in individuals at above average genetic risk for psychosis. *Psychological Medicine, 41*(11), 2305–2315. https://doi.org/10.1017/S0033291711000602.

Collip, D., Wigman, J. T. W., Myin-Germeys, I., Jacobs, N., Derom, C., Thiery, E., Wichers, M., & van Os, J. (2013). From epidemiology to daily life: Linking daily life stress reactivity to persistence of

psychotic experiences in a longitudinal general population study. *PLoS One, 8*(4), e62688. https://doi.org/10.1371/journal.pone.0062688.

Courvoisier, D. S., Eid, M., & Lischetzke, T. (2012). Compliance to a cell phone-based ecological momentary assessment study: The effect of time and personality characteristics. *Psychological Assessment, 24*(3), 713–720. https://doi.org/10.1037/a0026733.

Csikszentmihalyi, M., Larson, R., & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of Youth and Adolescence, 6*, 281–294. https://doi.org/10.1007/BF02138940.

Delespaul, P. (1995). *Assessing schizophrenia in daily life: The experience sampling method*. Maastricht, the Netherlands: Universitaire Pers Maastricht.

Edwards, C. J., Cella, M., Tarrier, N., & Wykes, T. (2016). The optimisation of experience sampling protocols in people with schizophrenia. *Psychiatry Research, 244*, 289–293. https://doi.org/10.1016/j.psychres.2016.07.048.

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2020). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment,* 1073191120957102. https://doi.org/10.1177/1073191120957102.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression*. Sage Publications Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Hufford, M. R. (2007). Special methdological challenges and opportunities in ecological momentary assessment. In A. A. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 54–75). Oxford University Press.

Jones, K. K., Zenk, S. N., McDonald, A., & Corte, C. (2016). Experiences of African-American women with smartphone-based ecological momentary assessment. *Public Health Nursing, 33*(4), 371–380. https://doi.org/10.1111/phn.12239.

Larson, R., & Csikszentmihalyi, M. (1978). Experiential correlates of time alone in adolescence. *Journal of Personality, 46*(4), 677–693. https://doi.org/10.1111/j.1467-6494.1978.tb00191.x.

Lataster, T., Collip, D., Lardinois, M., Van Os, J., & Myin-Germeys, I. (2010). Evidence for a familial correlation between increased reactivity to stress and positive psychotic symptoms. *Acta Psychiatrica Scandinavica, 122*(5), 395–404. https://doi.org/10.1111/j.1600-0447.2010.01566.x.

Lataster, T., Valmaggia, L., Lardinois, M., Van Os, J., & Myin-Germeys, I. (2013). Increased stress reactivity: a mechanism specifically associated with the positive symptoms of psychotic disorder. *Psychological Medicine, 43*(7), 1389–1400. https://doi.org/10.1017/S0033291712002279.

Litt, M. D., Cooney, N. L., & Morse, P. (1998). Ecological momentary assessment (EMA) with treated alcoholics: Methodological problems and potential solutions. *Health Psychology, 17*(1), 48–52. https://doi.org/10.1037/0278-6133.17.1.48.

Messiah, A., Grondin, O., & Encrenaz, G. (2011). Factors associated with missing data in an experience sampling investigation of substance use determinants. *Drug and Alcohol Dependence, 114*(2–3), 153–158. https://doi.org/10.1016/j.drugalcdep.2010.09.016.

Myin-Germeys, I., & Van Os, J. (2007). Stress-reactivity in psychosis: Evidence for an affective pathway to psychosis. *Clinical Psychology Review, 27*(4), 409–424. https://doi.org/10.1016/j.cpr.2006.09.005.

Myin-Germeys, I., Van Os, J., Schwartz, J. E., Stone, A. A., & Delespaul, P. A. (2001). Emotional reactivity to daily life stress in psychosis. *Archives of General Psychiatry, 58*(12), 1137–1144. https://doi.org/10.1001/archpsyc.58.12.1137.

Myin-Germeys, I., Peeters, F., Havermans, R., & Nicolson, N. a, DeVries, M. W., Delespaul, P., & Van Os, J. (2003). Emotional reactivity to daily life stress in psychosis and affective disorder: An experience sampling study. *Acta Psychiatrica Scandinavica, 107*(2), 124–131. https://doi.org/10.1034/j.1600-0447.2003.02025.x.

Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & Van Os, J. (2009). Experience sampling research in psychopathology: Opening the black box of daily life. *Psychological Medicine, 39*(9), 1533–1547. https://doi.org/10.1017/S0033291708004947.

Myin-Germeys, I., Klippel, A., Steinhart, H., & Reininghaus, U. (2016). Ecological momentary interventions in psychiatry. *Current Opinion in Psychiatry, 29*(4), 258–263. https://doi.org/10.1097/YCO.0000000000000255.

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.

Pawlik, K., & Buse, L. (1982). Computer-based behavior registration in the field: Description and first psychometric evaluation of a new recording method. *Zeitschrift fur Differentielle und Diagnostische Psychologie, 3*(2), 101–118.

R Development Core Team. (2016). R: A language and environment for statistical computing. *R Foundation for Statistical Computing Vienna Austria., 95*, 389–393. https://doi.org/10.1038/sj.hdy.6800737.

Rintala, A., 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. https://doi.org/10.1037/pas0000662.

Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2020). Momentary predictors of compliance in studies using the experience sampling method. *Psychiatry Research, 286*, 112896. https://doi.org/10.1016/j.psychres.2020.112896.

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

Solhan, M. B., Trull, T. J., Jahng, S., & Wood, P. K. (2009). Clinical assessment of affective instability: Compating EMA indices, questionnaire reports, and retrospective recall. *Psychological Assessment, 21*(3), 425–436. https://doi.org/10.1037/a0016869.

Stone, A. A., Broderick, J. E., Schwartz, J. E., Shiffman, S., Litcher-Kelly, L., & Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: Reactivity, compliance, and patient satisfaction. *Pain, 104*(1–2), 343–351. https://doi.org/10.1016/S0304-3959(03)00040-X.

Trull, T. J., & Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: Introduction to the special section. *Psychological Assessment, 21*(4), 457–462. https://doi.org/10.1037/a0017653.

Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and retention with the experience sampling method over the continuum of severe mental disorders: Meta-analysis and recommendations. *Journal of Medical Internet Research, 21*(2), e11398. https://doi.org/10.2196/14475.

Van Eck, M., Berkhof, H., Nicolson, N., & Sulon, J. (1996). The effects of perceived stress, traits, mood states, and stressful daily events on salivary cortisol. *Psychosomatic Medicine, 58*(5), 447–458. https://doi.org/10.1097/00006842-199609000-00007.

Van Genugten, C. R., Schuurmans, J., Lamers, F., & Riese, H. (2020). Experienced burden of and adherence to smartphone-based ecological momentary assessment in persons with affective disorders. *Journal of Clinical Medicine, 9*(2), 322. https://doi.org/10.3390/jcm9020322.

Van Roekel, E., Keijsers, L., & Chung, J. M. (2019). A review of current ambulatory assessment studies in adolescent samples and practical

recommendations. *Journal of Research on Adolescence, 29*(3), 560–577. https://doi.org/10.1111/jora.12471.

Verhagen, S. J. W., Hasmi, L., Drukker, M., van Os, J., & Delespaul, P. A. E. G. (2016). Use of the experience sampling method in the context of clinical trials. *Evidence-Based Mental Health, 19*(3), 86–89. https://doi.org/10.1136/ebmental-2016-102418.

Wen, C. K. F., Schneider, S., Stone, A. A., & Spruijt-Metz, D. (2017). Compliance with mobile ecological momentary assessment protocols in children and adolescents: A systematic review and meta-analysis. *Journal of Medical Internet Research, 19*(4), e132. https://doi.org/10.2196/jmir.6641.

Whalen, C. K., Jamner, L. D., Henker, B., Delfino, R. J., & Lozano, J. M. (2002). The ADHD spectrum and everyday life: Experience sampling of adolescent moods, activities, smoking, and drinking. *Child Development, 73*(1), 209–227. https://doi.org/10.1111/1467-8624.00401.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.