



## OPEN ACCESS

## EDITED BY

Gudberg K. Jonsson,  
University of Iceland, Iceland

## REVIEWED BY

Johannes Thrul,  
Johns Hopkins University, United States  
Aditya Ponnada,  
MongoDB, United States  
Mandani Ntekouli,  
Maastricht University, Netherlands

## \*CORRESPONDENCE

Stefan Schneider  
✉ schneids@usc.edu

RECEIVED 28 September 2024

ACCEPTED 06 December 2024

PUBLISHED 03 January 2025


## CITATION

Schneider S, Toledo MJ, Junghaenel DU,  
Smyth JM, Lee P-J, Goldstein S, Pomeroy O  
and Stone AA (2025) Do delayed responses  
introduce bias in ecological momentary  
assessment? Evidence from comparisons  
between self-reported and objective  
physical activity.  
*Front. Psychol.* 15:1503411.  
doi: 10.3389/fpsyg.2024.1503411

## COPYRIGHT

© 2025 Schneider, Toledo, Junghaenel,  
Smyth, Lee, Goldstein, Pomeroy and Stone.  
This is an open-access article distributed  
under the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other forums  
is permitted, provided the original author(s)  
and the copyright owner(s) are credited and  
that the original publication in this journal is  
cited, in accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Do delayed responses introduce bias in ecological momentary assessment? Evidence from comparisons between self-reported and objective physical activity

Stefan Schneider<sup>1,2,3\*</sup>, Meynard J. Toledo<sup>1</sup>,  
Doerte U. Junghaenel<sup>1,2,3</sup>, Joshua M. Smyth <sup>4</sup>,  
Pey-Juan Lee<sup>1</sup>, Sarah Goldstein<sup>1</sup>, Olivia Pomeroy<sup>1</sup> and  
Arthur A. Stone<sup>1,2</sup>

<sup>1</sup>Dornsife Center for Self-Report Science and Center for Economic and Social Research, University of Southern California, Los Angeles, CA, United States, <sup>2</sup>Department of Psychology, University of Southern California, Los Angeles, CA, United States, <sup>3</sup>Leonard Davis School of Gerontology, University of Southern California, Los Angeles, CA, United States, <sup>4</sup>Department of Psychology, The Ohio State University, Columbus, OH, United States

**Introduction:** Delayed responses are a common yet often overlooked aspect of participant compliance in ecological momentary assessment (EMA) research. This study investigated whether response delays introduce selection bias in the moments captured by EMA.

**Methods:** Participants ( $n = 339$ ) self-reported their physical activity behaviors using EMA five times a day over 7 days while wearing a continuous physical activity monitor. The continuous activity monitor data provided an objective reference value to evaluate potential biases in delayed EMA self-reports.

**Results:** Results showed that participants were significantly more likely to delay EMA responses when they were prompted during higher levels of physical activity, and they subsequently reduced their activity levels, postponing their response until they were in a significantly less active state. There was no significant evidence that response delays systematically biased the levels of EMA reported activities, although delayed responses were associated with significantly more random errors in EMA reports (with small effect sizes).

**Discussion:** The results suggest that respondents self-select the moments for answering EMA surveys based on their current activity levels, but brief response delays do not translate into marked reductions in the quality of EMA data.

## KEYWORDS

ecological momentary assessment (EMA), ambulatory assessment (AA), physical activity, response delay, compliance, experience sampling method (ESM)

## Introduction

The use of ecological momentary assessment (EMA) methods has substantially risen in popularity in psychological research over the past decades (Stone et al., 2023a). In EMA studies, participants are typically asked to complete brief surveys about their current behaviors, environments, and experiences, several times per day and typically over multiple days (Stone and Shiffman, 1994). Even though this assessment strategy yields rich and granular information about the ebb and flow of people's behaviors and experiences in everyday life, it is also demanding for participants to respond to all prompts in a timely manner. Over the course of a study, they are expected to briefly but repeatedly interrupt their ongoing daily activities to answer EMA questions when prompted to do so. Moreover, as many EMA studies intentionally signal respondents at (stratified) random times of the day, participants often cannot anticipate the timing of the assessments, and they are asked to respond regardless of whether or not this is currently convenient to them (Shiffman et al., 2008).

In view of this respondent burden, ensuring that participants comply reasonably well with the protocol has been a salient concern in EMA research for many years. To make reliable and valid conclusions from EMA data, it is desirable that the momentary data collected are a representative sample of respondents' experiences and behaviors in their daily lives (Shiffman et al., 2008). If participants do not adhere to the study protocol and self-select the moments they report on, selection bias (or sampling bias) occurs, because not all situations or experiences have an equal chance of being measured. The most overt form of protocol noncompliance is when participants show poor momentary response rates and miss EMA prompts. Given the perils associated with (especially nonrandom) missing data (Graham, 2012), a sizeable body of EMA research has examined the correlates and consequences of missing data due to skipped EMA assessments (see, for example, Jones et al., 2019; Ono et al., 2019; Vachon et al., 2019; Wrzus and Neubauer, 2023).

## Delayed responding to EMA prompts

Apart from skipping prompts entirely, participants may also engage in more subtle forms of noncompliance, such as delaying responses to EMA prompts (Eisele et al., 2021). In fact, one routinely employed strategy to reduce the amount of missing data due to skipped EMA prompts is to allow participants to temporarily "snooze" prompts or postpone their responses if the signal happens at an inopportune moment; that is, to complete the assessment within a certain time window *after* an EMA prompt occurred (Christensen et al., 2003; Stone et al., 2023a). In a typical EMA study, the delayed assessment is meant to retrospectively collect information about what was happening at the initial prompt. In this manner, the fidelity of the EMA sampling scheme is in theory preserved, because information is collected "as if" there was no delay in the response. If the targeted information is not recovered in delayed reports because respondents do not adequately recall the information from the initial prompt and/or report what was happening after the prompt, then delayed responses may not maintain sampling fidelity.

In most studies, researchers select a maximum delay period after which prompts can either not be answered or after which answers are later discarded from the data and are treated as if they were missed. The duration of the maximum allowed response delays is often constrained to less than 20–30 min (Christensen et al., 2003), even though studies have allowed delays ranging from as little as 1 min (Businelle et al., 2016) to more than 1 h after the occurrence of an EMA prompt (Bekman et al., 2013). Few articles provide an explicit rationale for the selected maximum allowable delay or document the response delay observed in a study (De Vries et al., 2021).

Even though response delays are widely employed and accepted in EMA studies, the potential implications of delayed responding have found limited attention in methodological EMA research (Eisele et al., 2021; Stone et al., 2023a). The concern addressed in this study is whether delayed responding creates a selection bias in the moments captured with EMA. When participants are allowed to postpone their responses, they can (to some degree) self-select the moments during which they complete a given EMA prompt. There are reasons to assume that this may in turn yield a biased representation of people's daily life experiences. For example, respondents may be more likely to delay prompts when they are currently very active (e.g., when they are exercising), or experiencing strong emotions (e.g., when they are arguing with another person), and they may delay the response until they have returned to a more calm or "normal" state. If respondents report on this more calm or "normal" present state rather than the state at the time of the prompt (either because they follow explicit EMA instructions to report on their experiences "right now" or because they are unable to recall their experiences when EMA items ask them to report what happened "at the prompt"), this could lead to an underrepresentation of very active or emotionally intense moments collected with EMA (Eisele et al., 2021). If this was the case, it could affect measures of frequency, intensity, or duration of behaviors and experiences across individuals in between-person analyses, and it could also distort (e.g., reduce) the magnitude of observed fluctuations within individuals in within-person analyses. To date, little is known about whether delayed responses are merely a harmless departure from the optimal study protocol or if they indeed introduce significant and meaningful selection bias in EMA (Stone et al., 2023a).

## Previous research on EMA response delays

The few studies that have been conducted in this area roughly fall into two categories. The first type of study has focused on identifying potential predictors of EMA response delays. Sarker et al. (2014) examined physiological sensor data collected from 30 university students and found that participants were least likely to delay responding to a prompt when they were walking outside, and most likely to delay at work and when they were driving. Similarly, Boukhechba et al. (2018) found that combinations of passively recorded contextual variables (time of day, location, and phone movement) and participant-phone interactions predicted the length of response delays in 65 university students with moderate accuracy. These results support the idea that momentary behaviors and environmental characteristics are systematically

associated with delayed responding in EMA. A strength of these studies is that they analyzed variables that were passively recorded in real time at the time of the delays. At the same time, because the EMA reports themselves were not analyzed, these results do not answer the question whether response delays in fact introduced biases in participants' self-reported momentary behaviors and experiences.

The second type of study has examined differences in self-reported experiences between prompts to which participants responded immediately and prompts for which they delayed the response. Affleck et al. (1998) compared fibromyalgia patients' reports of momentary pain, mood, and fatigue between random selections of 100 delayed and 100 immediate responses that were matched by time of day and day of week; they did not find significant differences for either pain, mood, or fatigue reports. In contrast, a large secondary data analysis of over 1,500 individuals from nine different paper-and-pencil EMA studies (Eisele et al., 2021) showed that participants tended to rate their positive affect as significantly higher and negative affect as significantly lower when they had delayed their responses, as compared to immediate EMA responses. Effect sizes were small yet consistent across different (nonclinical and clinical) samples (Eisele et al., 2021). Even though these results suggest that delayed momentary self-reports may be different (e.g., more positive) on average than immediate reports, they do not allow firm conclusions about the extent to which delayed responding introduces selection bias in EMA. If, for example, participants tend to delay responding in situations that are associated with more extreme mood states (e.g., high negative or low positive affect) and respond at their convenience when their mood levels have normalized, the sampled mood levels may not differ between immediate and delayed responses despite (or precisely because of) self-selection. The counterfactual question "what value would a person who delayed a response have provided had they *not* delayed the response" is not directly addressed by comparing EMA ratings between delayed and immediate responses.

## The present study

Taken together, prior literature provides a fragmented picture of the impact of response delays on the representativeness of data collected and does not offer firm evidence-based guidance on the use of response delays in EMA studies. The goal of the present study was to conduct a rigorous analysis addressing the question whether response delays introduce selection biases in EMA. To this end, we examined both passive real-time recordings and momentary self-reports of the same construct, that is, we combined the two approaches that previous studies used to examine response delays in the same analysis. More specifically, we examined pre-existing data from an EMA study in which physical activity behaviors were assessed both continuously with accelerometry and with EMA self-reports over the same time period. To illustrate why it is uniquely beneficial to be able to draw on passive recordings and EMA reports of the same construct in combination, one may think of delayed EMA responding as a process of replacement ("imputation") of missing values: participants initially omit (i.e., miss) a response when prompted, and then replace (i.e., impute) the omitted response with a new value at a later point in time. A critical

question is how well the new (observed) value reflects the value of the originally omitted response. Selection bias occurs when the omitted EMA responses are replaced with delayed EMA responses that differ from the omitted ones. By having passive recordings of the same construct assessed with EMA available for the exact time of the omitted EMA response, we were able to use these passive recordings as a reference point or "benchmark" to evaluate how closely the delayed EMA responses reflected the values of the originally omitted responses.

The specific construct used to investigate whether response delays are associated with selection bias targeted people's physical activity behaviors. Admittedly, physical activity is merely one type of behavior and, as must be the case, is not necessarily representative of the broad spectrum of behaviors and experiences commonly assessed in clinical EMA studies (Shiffman et al., 2008). However, physical activity patterns in daily life are associated with dynamic changes in many constructs relevant to clinical psychology, including everyday changes in affect (Wen et al., 2018), stress (Do et al., 2021), and cognitive function (Zlatař et al., 2022). Moreover, as noted above, to study potential biases associated with temporarily omitted (i.e., delayed) EMA reports, we required a strategy that allowed comparing subjective EMA reports with continuous measurements of their objective counterparts; the construct of physical activity behaviors lends itself ideally to parallel assessments with EMA and with passive recordings via accelerometry (Stinson et al., 2022).

Several research questions were addressed. The first two questions asked whether objective physical activity levels measured with accelerometry (a) *predicted* the occurrence of response delays, and (b) whether these activity levels *changed* over the course of a response delay; together, this addresses the question whether people tend to systematically self-select the moments at which they complete EMA assessments. The third question asked whether the self-selection of moments of an EMA response translates to biases in *self-reported* activity levels in EMA.

### Question 1: do momentary physical activity levels predict delayed EMA responding?

Our first analysis focused on objective (i.e., passively recorded) physical activity levels at the time (i.e., during the minute) of a given EMA prompt. Our question was whether physical activity levels at the time of the prompt are systematically associated with the likelihood of delaying a response. We hypothesized that individuals will be less likely to delay EMA prompts when they are more sedentary and more likely to delay EMA prompts when they are more physically active at the time of a prompt, because being currently physically active may make it more difficult and less convenient to immediately respond to an EMA prompt (Dunton, 2017).

### Question 2: is delayed responding associated with changes in physical activity levels?

Our second question was whether people's objective physical activity levels *changed* over the course of a response delay, that is, from the time of the EMA prompt to the time of the actual response. Selection biases in EMA are arguably more likely to occur if respondents are objectively in a systematically different state when they respond after a delay compared to their state when they were initially prompted. We hypothesized that individuals will tend

to respond after an observed decrease in their physical activity (i.e., from elevated activity levels to a more resting state from the time of a prompt to the time of the delayed response).

### Question 3: how well do delayed self-reports of physical activity reflect respondents' objective physical activity levels at the time of the initial EMA prompt?

Our third question was whether the self-selection of moments of an EMA response translates into actual biases in EMA self-reports. Even if the moments at which individuals choose to respond to EMA prompts are self-selected based on current activity levels, this does not necessarily imply that delayed EMA ratings differ from the answers people would have given had they responded right away. To address this question, we compared people's self-reported physical activity levels with their objective physical activities that were passively recorded at the time of the prompt. This allowed us to examine if delayed self-reports reflected people's objective physical activity levels at the time of the prompt less well than self-reports that were not delayed. More specifically, we examined if delayed reports introduce (a) systematic biases and (b) random errors in people's EMA self-reports.

#### Question 3a: are response delays associated with systematic biases in the levels of self-reported physical activity?

When individuals change their physical activity levels from the time of the prompt to the time of a delayed response, their self-reported physical activity levels may be affected by response delays, as well, given that self-reports are impacted by people's present states (Schneider and Stone, 2016). We hypothesized that response delays would be associated with systematically lower self-reported physical activity levels compared to what would be expected based on objective physical activity levels recorded at the time of the prompt.

#### Question 3b: are response delays associated with more random errors in people's self-reported momentary physical activities?

Apart from systematic shifts in the levels of people's self-reports, delaying an EMA response to a later time point may reduce their ability to precisely report their behaviors at the time of the prompt. Memory for momentary experiences and behaviors fades quickly over time (Conner and Barrett, 2012; Walentynowicz et al., 2018), and people may not be fully able to recall their physical activities accurately even with a delay of just a few minutes after an EMA prompt. Accordingly, we hypothesized that self-reported physical activities are less strongly correlated with objective physical activity levels recorded at the time of the prompt when people delayed their EMA responses compared to when EMA responding was not delayed.

objectively recorded physical activity levels. Study participants were recruited from the Understanding America Study (UAS) (Alattar et al., 2018), a probability-based internet panel study of about 13,000 US residents. Members are recruited to be part of the UAS through address-based sampling, and individuals without prior internet access receive a tablet and broadband access (excluding smartphones). This ensures broad coverage in populations typically underrepresented in opt-in or volunteer online panels. UAS panelists were eligible to participate in the EMA study if they (1) were fluent in English; (2) were 18 years of age or older; (3) had no hearing impairments, (4) had no vision impairments that could not be corrected with contact lenses or glasses; (5) did not work a night shift; (6) had stable access to email; (7) had daily access to Wi-Fi; (8) were not on bed-rest; and (9) did not require any mobility devices to be able to move around. Panelists were contacted via email to be invited into the study and to complete the eligibility screening survey. Of 1,363 UAS panelists who were contacted and assessed for eligibility, 1,021 met the eligibility criteria. Of those, 407 provided consent and were enrolled into the study, and 359 completed the study protocol. The study was approved by the Biomedical Research Alliance of New York Institutional Review Board (#22-183-1044). Participants provided electronic informed consent for participation.

## Study procedures

The study included a 7-day EMA protocol that was executed remotely; participants were mailed a study package to their home that contained the study devices (a smartphone with charger and an activity monitor with medical tape and alcohol swabs) and mailing materials for returning the package after completing the EMA study. The day before starting the protocol, participants were instructed to watch a 20-min training video that explained how to put on and wear the activity monitor, and that guided participants through all aspects of EMA survey completion (how to interact with the device and app, the questions presented, and how to enter their responses). Starting on the next day, participants received five EMA prompts per day for 7 days while continuously wearing their physical activity monitor. Research staff checked the number of completed EMA prompts daily and contacted participants over the telephone and through email if there were suspected technological problems. Participants received email reminders about completing the study activities on a daily basis. Upon study completion, participants returned their study package using the provided mailing materials and prepaid return label. Participants were compensated up to \$200 for study participation, which was prorated based on their level of overall protocol completion: participants received \$15 for completing the training, \$10 for completing the baseline questionnaire, \$2 for each EMA survey completed, \$10 for each day of wearing the activity monitor, \$10 for completing a final questionnaire, and \$25 for returning the study devices at the end of the protocol.

## Materials and methods

### Participant sample

The data analyzed for this study came from a larger EMA protocol examining relationships between self-reported and

### Accelerometry-based physical activity measurement

The activPAL<sup>TM</sup> activity monitor was used to objectively measure participants' physical activity levels over the study period.



Participants were asked to wear the monitor for seven consecutive days, 24 h a day, except during water-based activities. Unlike hip or wrist-worn accelerometry-based activity monitors, the activPAL is worn on the anterior midline of the thigh. This allows to detect the orientation of the participant's thigh and to discriminate different postures (upright versus seated or lying down). Upon return of the study device, the raw accelerometer data was downloaded and processed using activPAL's software (PALanalysis v9.1.0.102). Using the software's proprietary algorithm (CREA v1.3 algorithm) the raw accelerometry data was processed into events of (a) sedentary (i.e., sitting or lying down), (b) standing, and (c) stepping behaviors (i.e., step counts). Numerous validation studies support the accuracy of the device and algorithm (e.g., Bassett et al., 2014; Lyden et al., 2017). We further used the cadence of momentary stepping activities to distinguish times spent in (d) light physical activity (LPA; cadence < 100 steps/min), (e) moderate physical activity (MPA; cadence 100–125 steps/min), and (f) vigorous physical activity (VPA; cadence > 125 steps/min) levels (Tudor-Locke and Rowe, 2012).

## EMA data collection

As part of the larger study protocol, EMA items were administered using two different reporting periods such that the EMA questions asked participants about their behaviors and experiences either during “the 5 min before the prompt” or during “the 2 h before the prompt”; the reporting period was assigned to participants in a between-person randomized design. EMA prompts were delivered using a stratified random sampling scheme with five prompts that could occur between 6 a.m. and 10 p.m. in the participant's time zone, and which was stratified such that the sampled reporting periods covered by the EMA questions were at least 15 min apart. For respondents who were asked to report about the 5 min before the prompt, this was accomplished by dividing the waking day into five equal time bins and randomly selecting prompt times within each of these bins (provided a minimum of 20 min between two prompts). For respondents who were asked to report about the 2 h before the prompt, the waking day was divided into equal time bins with a duration slightly longer than 2 h each. From these time bins, a subset of five bins were randomly selected each day for assessment. EMA questions were administered using the movisensXS app on study provided smartphones (Motorola G Power 2021).

Self-reported physical activity levels were assessed with five EMA items asking participants how many minutes they spent (1) sitting or lying down (i.e., sedentary), (2) standing, (3) in LPA, (4) in MPA, and (5) in VPA during the targeted time (i.e., 5 min or 2 h) before the EMA prompt. The EMA questions were designed to closely mirror the physical activity variables derived from the activPAL activity monitor, while also aligning with items used in prior research on the validity of EMA reports (Stinson et al., 2022). Answers were given by entering the number of minutes using an open-ended numeric response format. Two additional EMA items asked participants how “sedentary” and how “physically active” they were during the targeted time (5 min or 2 h) before the prompt, with answers provided using a 5-point verbal rating scale (not at all – extremely). Respondents reported the sleep and wake time for

the past night during the first EMA prompt of each day. EMA items on affect, pain, and fatigue were also administered but were not analyzed here.

Participants were allowed to delay their responses to each EMA prompt for up to 15 min. After an initial alarm that lasted for 1 min, reminder alarms lasting 1 min each were administered by the app 5 and 10 min after the first alarm. The prompt expired 15 min after the beginning of the initial alarm and the participant was unable to complete the EMA survey.

## Analysis strategy

Study participants were included in the present analyses if they provided accelerometer data for all 7 study days and for at least 70% of their waking hours during the week. This excluded 20 of the 359 participants, with a resulting analysis sample size of  $n = 339$ ; of those,  $n = 173$  answered EMA questions about the 5 min before the prompt and  $n = 166$  reported on the 2 h before the EMA prompt.

Response latencies for all EMA prompts were calculated as the difference (in seconds) between the time the prompt was delivered by the app and the time a respondent started the survey. We categorized responses as “delayed” if a participant started the survey more than 1 min after the EMA prompt, that is, following the end of the initial alarm. Conversely, responses were coded as “immediate” if the survey was initiated within 1 min of the EMA prompt, during the period when the initial alarm was still active. This categorization is akin to answering a phone call while it is still ringing.

Our analyses of response delays considered only EMA prompts that were administered during participants' waking hours. Periods of sleep and accelerometer non-wear time were excluded from the analyses. Sleep periods were determined using both passive sleep detection from accelerometer data (Courtney et al., 2021) and EMA self-reported sleep and wake times; the two sources of information were combined using a weighted average with an algorithm proposed by Thurman et al. (2018).

## Analyses for question 1: do respondents' activity levels predict response delays?

Multilevel logistic regression models were used to examine whether objective physical activities recorded by accelerometry during the minute of the EMA prompt predicted the occurrence of delayed EMA responding. The timeliness of an EMA response (immediate versus delayed) served as binary outcome variable. To evaluate the effects of physical activity behaviors at specific intensity levels, we entered the amount of time sedentary (i.e., lying/sitting), time standing, time in LPA, MPA, and VPA individually as predictor variables in separate regression models. In addition, respondents' step count during the minute of the EMA prompt was examined as a summary measure of overall physical activity intensity and was entered as predictor in a separate model.

For each physical activity variable, the log odds of a delayed response  $\eta_{ij}$  for participant  $j$  at a given measurement occasion  $i$  were estimated as described by the following multilevel equation:

$$\text{Level 1: } \eta_{ij} = \log\left(\frac{\text{delay}_{ij}}{1 - \text{delay}_{ij}}\right) = \beta_{0j} + \beta_{1j}PA_{ij}^{(WC)}$$

$$\text{Level 2 : } \beta_{0j} = \gamma_{00} + \gamma_{01}PA_j^{(M)} + u_{0j} \quad (1)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

We distinguished within-person and between-person and effects of physical activity. At level 1, the within-person level, the log odds of delaying a response were predicted from within-person centered physical activity variables  $PA_j^{(WC)}$ ; this addresses the question whether individuals are more likely to delay an EMA response when they are more physically active (or more sedentary) than they usually are. At level 2, the between-person level, individual differences in the propensity to delay EMA responses over the course of the study were predicted from person mean levels of the physical activity variables  $PA_j^{(M)}$ ; this addresses the question whether respondents who are generally more active (or more sedentary) are more likely to delay EMA responses. Also at level 2, the models allowed for residual variance in the random intercepts  $\beta_{0j}$  (individual differences in the propensity of response delays not explained by physical activity) and for between-person variance in the regression slopes  $\beta_{1j}$  (individual differences in the within-person effect of physical activity on response delays).

To facilitate interpretation of the magnitude of between-person variance in the regression slopes  $\beta_{1j}$ , we computed the coefficient of variation (CV) as a measure of relative variability. The CV was calculated as the ratio of the standard deviation in regression slopes to the mean regression slope (i.e., the fixed effect,  $\gamma_{10}$ ). A CV of 1.0 indicates that the regression slopes for most individuals (approximately 68%, assuming normally distributed random effect variances) range between 0 (no effect) and two times the average slope, suggesting substantial heterogeneity in effects.

### Analyses for question 2: is delayed responding associated with changes in activity?

To investigate whether respondents' objective activity levels changed from the time of the initial alarm to the time of a delayed response to an EMA prompt, we analyzed the accelerometry data recorded for two time points: (a) the time point (i.e., minute) of the EMA prompt, and (b) the time point (i.e., minute) the participant responded, among those EMA responses that were delayed. Linear multilevel models for change between the two time points were estimated separately for each of the physical activity indicators (amount of time sedentary, standing, time in LPA, MPA, VPA, and step count) as described by the following multilevel equation:

$$\text{Level 1 : } PA_{hij}^{OBJ} = \beta_{0j} + \beta_{1j}TIMEPOINT_{hij} + r_{hij},$$

where  $r_{hij} \sim N(0, \Sigma)$

$$\text{Level 2 : } \beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

The accelerometry-derived measurements taken at the time of the prompt and the time of the response simultaneously served as multivariate (i.e., repeated-measures) dependent variable.

$PA_{hij}^{OBJ}$  represents the objective physical activity measurement for individual  $j$  and measurement occasion  $i$  at time point (time of prompt or response)  $h$  (Snijders and Bosker, 1999). The dichotomous predictor variable *TIMEPOINT* was coded as 0 for measurements taken at the time of the prompt and as 1 for measurements taken at the time of the delayed response. Accordingly, the level 1 intercept  $\beta_{0j}$  represents a given respondents' physical activity level at the time of the prompt (on average across measurement occasions), and the regression slope  $\beta_{1j}$  represents the respondent's change in activity from the alarm to the time of the delayed response (on average across measurement occasions). Random effects for the intercept  $\beta_{0j}$  and rate of change  $\beta_{1j}$ , modeled at level 2, allow for individual differences in both parameters.

To account for nonindependences of activity levels at the time of an EMA prompt and the time of the response within a given measurement occasion, the within-person residuals  $r_{hij}$  at level 1 were allowed to correlate through the error variance-covariance matrix  $\Sigma$ . Specifically, an unstructured covariance matrix was used to freely estimate the residual within-person activity variances at the time of the prompt and time of the response, as well as their covariance, across measurement occasions.

### Analyses for question 3: how well do delayed EMA reports reflect objective activity levels at the time of the prompt?

As described above, participants for this study were randomized to receive one of two different versions of the same EMA items that either asked about "the 5 min before the prompt" or "the 2 h before the prompt". These subsamples were analyzed separately in the analyses comparing self-reported and objective physical activity levels described below. Because EMA typically employs brief reporting periods, we present primary results based on the subsample of respondents who received the 5-min version of the EMA items,  $n = 173$ . Secondary results from the subsample of respondents receiving the 2-h version of the items ( $n = 166$ ) are shown in [Supplementary material](#).

### Analysis for question 3a: are response delays associated with systematic biases in the levels of self-reported physical activity?

To examine whether response delays introduced systematic bias in self-reported activity levels, we compared respondents' answers on each of the five EMA items (i.e., amount of time sedentary, standing, time in LPA, MPA, and VPA) with respondents' objective physical activity levels for corresponding variables recorded during the 5 min<sup>1</sup> before each prompt, and examined whether any activity level differences between these two assessment types were moderated by response delays. In other words, we examined the difference between EMA self-reported and objective measurements as a function of response delay. Linear

<sup>1</sup> Objective physical activity levels recorded for the 5 min before the prompt were analyzed here in accordance with the time period the EMA reports in the primary analyses were supposed to capture. For the secondary analyses of respondents who received the 2-h version of EMA items, objective physical activity levels for the 120 min before the prompt were analyzed.

multilevel models were used for this purpose as described by the following equation:

$$\text{Level 1 : } PA_{hij} = \beta_{0j} + \beta_{1j}ASSESTYPE_{hij} + \beta_{2j}DELAY_{hij} + \beta_{3j}(ASSESTYPE*DELAY)_{hij} + r_{hij},$$

where  $r_{ij} \sim N(0, \Sigma)$

$$\begin{aligned} \text{Level 2 : } \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \\ \beta_{2j} &= \gamma_{20} + u_{2j} \\ \beta_{3j} &= \gamma_{30} + u_{3j} \end{aligned} \quad (3)$$

The self-reported and objective measurements of a physical activity variable (time sedentary, standing, LPA, MPA, and VPA; examined in separate models) simultaneously served as multivariate dependent variable. Thus,  $PA_{hij}$  represents the value of a physical activity variable for participant  $j$  at measurement occasion  $i$  and assessment type (i.e., self-reported or objective)  $h$ .

The dichotomous predictor variable  $ASSTYPE$  was coded as 0 for objective and as 1 for EMA self-reported assessments. The dichotomous predictor variable  $DELAY$  was coded as 0 for immediate and as 1 for delayed responses. Both predictors were within-person centered before computing the  $ASSTYPE \times DELAY$  interaction term so that they can be interpreted analogous to within-person main effects (Aiken et al., 1991; Yaremych et al., 2023). Accordingly, the regression slope  $\beta_{1j}$  represents the difference between EMA self-reported and objective activity levels, and the regression slope  $\beta_{2j}$  represents the difference in activity levels between prompts that were delayed versus not delayed. The interaction term  $\beta_{3j}$  compares self-reported and objective activity levels between immediate and delayed responses, testing the hypothesized effect. If the interaction term is significant, this indicates that the self-reported activity levels for delayed EMA responses differ from (i.e., are systematically higher or lower than) what would be expected had respondents answered the prompts immediately.

Level 2 random effects for the intercept  $\beta_{0j}$  and regression parameters  $\beta_{1j}$ ,  $\beta_{2j}$ , and  $\beta_{3j}$  allowed for individual differences in each of these parameters. To account for nonindependences of a person's self-reported and objectively recorded activity levels at the same measurement occasion, the within-person residuals  $r_{hij}$  at level 1 were allowed to correlate through an unstructured variance-covariance matrix  $\Sigma$ .

### Analysis for question 3b: are response delays associated with more random errors in self-reported physical activities?

Finally, to examine whether response delays introduced unsystematic ("random") errors in respondents' EMA reports, we examined whether the magnitude of associations between self-reported and objective activity levels was moderated by response

delays. The following multilevel model was estimated for this purpose:

$$\text{Level 1 : } PA_{ij}^{OBJ} = \beta_{0j} + \beta_{1j}PA_{ij}^{EMA} + \beta_{2j}DELAY_{ij} + \beta_{3j}(PA^{EMA} \times DELAY)_{ij} + r_{ij}$$

where  $r_{ij} \sim N(0, \sigma^2)$

$$\begin{aligned} \text{Level 2 : } \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \\ \beta_{2j} &= \gamma_{20} \\ \beta_{3j} &= \gamma_{30} \end{aligned} \quad (4)$$

Here,  $PA_{ij}^{OBJ}$  represents the objective physical activity variable (time sedentary, standing, LPA, MPA, and VPA; examined in separate models) recorded over the 5 min before the prompt for occasion  $i$  and person  $j$ . At level 1, it was regressed on the within-subject centered EMA self-report item for the same physical activity,  $PA_{ij}^{EMA}$ , a within-subject centered binary indicator for delayed versus immediate responses,  $DELAY_{ij}$ , and the EMA self-report by response delay interaction,  $PA_{ij}^{EMA} \times DELAY_{ij}$ . Accordingly, the regression slope  $\beta_{1j}$  represents the magnitude of the within-subject association between self-reported and objective physical activities, and the regression slope  $\beta_{2j}$  represents the difference in objective activity levels between prompts that were delayed versus not delayed. The interaction term  $\beta_{3j}$  indicates whether the magnitude of a within-subject association between self-reported and objective physical activities is moderated by delayed responses, testing the hypothesized effect. If the interaction is significant, this indicates that the association differs between responses that were delayed versus not delayed.

At level 2, the intercepts  $\beta_{1j}$  and regression slopes of the EMA self-reports  $\beta_{1j}$  were modeled as random effects. The remaining parameters  $\beta_{2j}$  and  $\beta_{3j}$  were not modeled as random effects because the between-person variances for these parameters were found to be zero in analyses after controlling for the fixed effect parameters.

In addition to the five EMA items that asked participants how many minutes they spent at different physical activity levels using an open-ended numeric response format, two EMA items that used a 5-point verbal rating scale format ("how sedentary were you?" and "how physically active were you?"; not at all – extremely) were also analyzed here to examine whether the strength of their association with corresponding objective physical activity variables (time sedentary; time in light or higher activity levels) was moderated by response delays.

This study's design and its analysis were not preregistered. We report fit statistics, including  $-2 \times \log$  likelihood, AIC, and BIC, alongside the parameter estimates to provide a transparent overview of model performance. These metrics quantify the balance between model fit and complexity. We did not conduct fit comparisons with more restrictive (error) variance structures; we freely estimated random effects for all model parameters unless this led to non-positive definite covariance matrices, to account for potential heterogeneity in the estimated effects across individuals. All data, analysis code, and research materials are available upon request from the first author. Data were analyzed using SAS version 9.4 (Cary, NC, USA).

## Sample size and statistical power

Statistical power in multilevel models depends on a series of factors, including sample size, number of repeated observations, the between- and within-person variance composition, and the magnitude of random effects variances (Bolger and Laurenceau, 2013). Because research questions 3a and 3b involved lower sample sizes, we estimated the smallest effect sizes detectable with a sample of  $n = 173$  and 30 observations per person (i.e., assuming 15% missing data for five daily EMA prompts over 7 days) for these two questions. Power calculations were conducted using Monte Carlo simulations consisting of 1,000 replications in *Mplus* version 8.10 (Muthén and Muthén, 2017). Based on prior EMA studies (Ono et al., 2019; Podsakoff et al., 2019) we assumed 15% delayed responses, a ratio of random intercept to within-person variances of 0.5/1, and a ratio of random intercept to random slope variances of 1/0.2 for all random effects. For question 3a, the analyzed sample provided 80% power to detect an effect size of 0.27 ( $\alpha = 0.05$ ) for the two-way within-person interaction (standardized difference in differences) between assessment type (EMA versus objective recordings) and response delay (immediate versus delayed responses). For question 3b, the sample provided 80% power to detect differences in standardized within-person regression coefficients between immediate and delayed responses as small as  $\beta = 0.035$  ( $\alpha = 0.05$ ), assuming a regression coefficient of 0.50 for the relationship between objectively recorded and EMA reported activities for immediate responses.

## Results

Demographic characteristics of the study sample are shown in Table 1. Participants' mean age was 48.83 years (SD = 14.91; range = 19–86 years). About half (51.78%) were female, four fifths (79.88%) were White, one seventh (14.79%) were Hispanic, and about three fourths (72.19%) were married. The majority (65.68%) had graduated college.

The distribution of response latencies in minutes for all individual EMA prompts is shown in Figure 1A. Out of 11,270 prompts, 9,125 (80.97%) were answered within 1 min of the prompt (i.e., before the initial alarm had ended) and were coded as immediate responses, 1,225 (10.87%) were answered more than 1 min after the prompt and were coded as delayed responses; the remaining 920 (8.16%) prompts were missed. Figure 1B shows the between-person distribution in the number (i.e., proportion of prompts) of delayed responses; the interquartile range of delayed responses was between 3.1% and 18.8% across participants. To descriptively examine the extent to which response delays were attributable to stable between-person or to momentary within-person factors, we calculated the intraclass correlation coefficient (ICC) for the occurrence of response delays.<sup>2</sup> The ICC was 0.25, indicating that about 25% of the variation in response delays was

TABLE 1 Demographic characteristics of the study sample ( $N = 339$ ).

	Frequency (%) or mean (SD)
Age (mean, SD)	48.84 (14.89)
Female	176 (51.92%)
<b>Race</b>	
White	271 (79.94%)
Black	22 (6.49%)
American Indian or Alaska Native	5 (1.47%)
Asian	17 (5.01%)
Native Hawaiian or Pacific Islander	3 (0.88%)
Mixed	21 (6.19%)
Hispanic	50 (14.75%)
<b>Education</b>	
Less than high school	8 (2.36%)
High school graduate	42 (12.39%)
Some college	66 (19.47%)
College graduate	35 (10.32%)
Bachelor's degree	98 (28.91%)
Master's degree or higher	90 (26.55%)
<b>Income<sup>a</sup></b>	
Less than \$30,000	45 (13.31%)
\$30,000–\$49,999	42 (12.43%)
\$50,000–\$99,999	105 (31.07%)
\$100,000–\$149,999	57 (16.86%)
\$150,000 or more	89 (26.33%)
<b>Marital status</b>	
Married	245 (72.27%)
Separated/divorced	32 (9.44%)
Widowed	7 (2.06%)
Never married	55 (16.22%)

<sup>a</sup>Income was not reported by one participant.

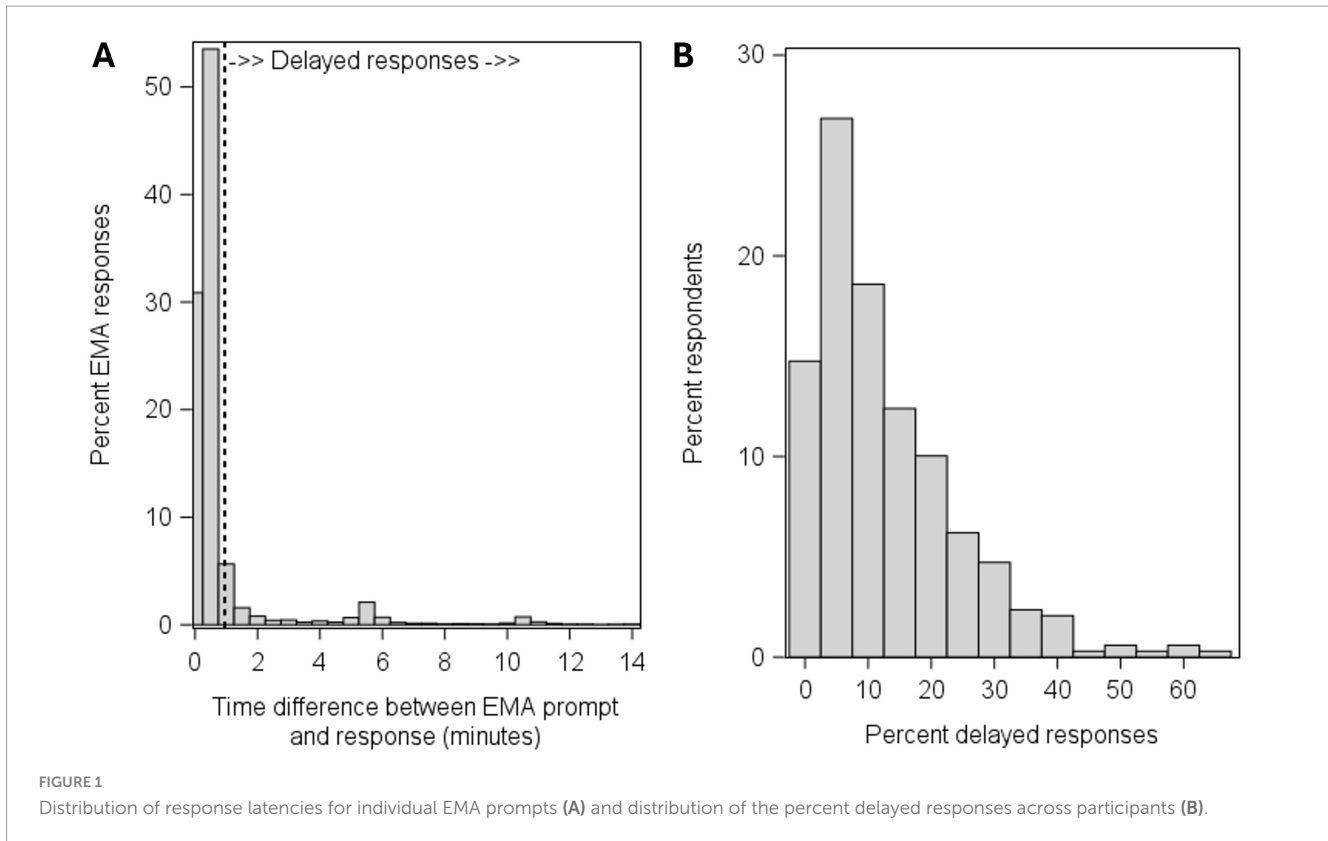
due to between-person differences, whereas the remaining 75% was attributable to moment- or situation-specific factors.

## Do objective activity levels predict EMA response delays?

As shown in Table 2, within-person effects of current physical activity levels predicting the occurrence of delayed responses were highly significant. The probability of delaying a response decreased to the extent that a respondent was more sedentary than usual, and increased when a respondent spent more time than usual standing, in LPA, MPA, and VPA. Specifically, the odds of delaying a response decreased by 10% for every 10 s a participant spent more time in sedentary states during the minute of the prompt, and increased by 9%, 24%, 13%, and 55% for every 10 s a participant spent more

<sup>2</sup> The intraclass correlation was calculated as  $\tau / (\tau + 3.29)$ , where  $\tau$  represents the estimated between-person variance of in a multilevel logistic regression model with no predictor variables, and 3.29 is the fixed within-person variance of a standard logistic distribution.





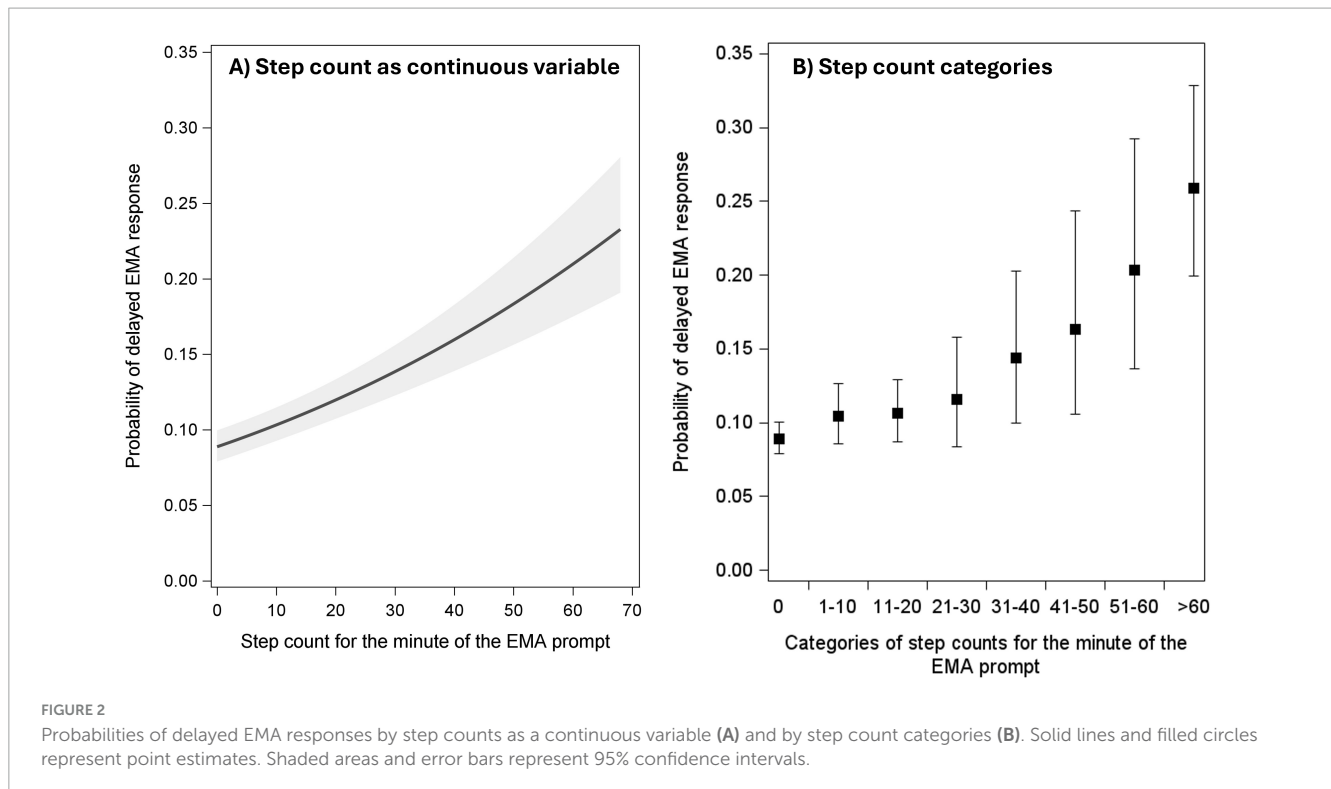
**TABLE 2** Results from multilevel logistic regression models predicting the occurrence of delayed responses from objective measures of physical activity recorded during the minute of the EMA prompt.

Parameter	Sedentary	Standing	LPA	MPA	VPA	Steps
<b>Fixed effects</b>						
Intercept	-1.694 (0.28)***	-2.361 (0.14)***	-2.449 (0.12)***	-2.260 (0.08)***	-2.201 (0.07)***	-2.458 (0.12)***
Within-person	-0.110 (0.01)*** OR = 0.90	0.087 (0.02)*** OR = 1.09	0.213 (0.03)*** OR = 1.24	0.118 (0.05)* OR = 1.13	0.438 (0.15)** OR = 1.55	0.128 (0.02)*** OR = 1.14
Between-person	-0.135 (0.07)* OR = 0.87	0.105 (0.09) OR = 1.11	0.520 (0.22)* OR = 1.68	0.631 (0.42) OR = 1.88	1.499 (1.57) OR = 4.48	0.325 (0.13)* OR = 1.38
<b>Variance components</b>						
Intercept	0.828	0.822	0.816	0.822	0.825	0.827
Within-person regression slope	0.011	0.014	0.004	0.000	0.000	0.001
Covariance	0.002	0.007	0.015	-0.001	0.000	0.010
<b>Fit statistics</b>						
-2 log likelihood	49,610.05	49,457.42	49,709.94	49,532.11	49,532.13	49,693.81
Generalized $\chi^2$	7,430.02	7,481.55	7,743.55	7,747.98	7,747.36	7,717.34
$\chi^2/df$	0.78	0.78	0.81	0.81	0.81	0.81

Standard errors are in parentheses. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ . Predictors are scaled in 10 s/min (for time sedentary, time standing, time in LPA, time in MVPA) or in 10 steps/min (for step count). LPA, light physical activity; MPA, moderate physical activity; VPA, vigorous physical activity; OR, odds ratio; df, degrees of freedom.

time standing, in LPA, MPA, and VPA, respectively. The effects of time spent sedentary and standing showed substantial variation between individuals: the between-person standard deviation of regression slopes was close to (CV = 0.95 for sedentary) or greater than (CV = 1.36 for standing) the mean regression slope for these variables. In contrast, the variability in the effects of LPA, MPA, VPA, and step counts was less pronounced (CV = 0.05–0.30).

Between-person effects of physical activity mirrored the pattern observed in the within-person effects: respondents who showed more sedentary behaviors in general were significantly less likely to delay EMA responses over the course of the study, and respondents who spent more time in physically active behaviors were more likely to delay responses, even though the latter effects were only statistically significant for time in LPA (see Table 2).



Corroborating these results, the probability of delayed responses also significantly increased with greater step counts at both within- and between-person levels of analysis (see Table 2). Figure 2A shows the linear effect of step counts as a continuous predictor variable. As it is also plausible that the effects of physical activity are nonlinear or discontinuous (e.g., physical activity levels might need to surpass a certain “threshold” to affect participants’ response behaviors), in secondary analyses, we divided the continuous step count variable into categories (0 steps, 1–10 steps, 11–20 steps, 21–30 steps, etc., during the minute of the prompt) and entered the resulting variable as a categorical predictor of response delays in a multilevel logistic regression model. Figure 2B shows the resulting estimated probability of delayed responses by step count category; the probability of delaying a response was 0.089 for when 0 steps were taken, and it monotonically increased for higher step count categories to a probability of 0.259 when 60 or more steps were taken during the minute of the EMA prompt.

We also explored the question whether physical activities occurring during the minutes *before* an EMA prompt would predict the likelihood of EMA response delays. To examine this, we examined the step counts recorded at increasing time lags (1 min, 2 min, etc., up to 5 min) before the EMA prompt. As shown in Table 3, when the step counts for each of these time lags were entered individually as predictors of response delays in separate multilevel logistic regressions, the effects remained significant for all time lags even though the effect sizes (odds ratios) decreased quickly from OR = 1.14 (for the minute of the prompt) to OR = 1.10 (1 min before the prompt) to OR = 1.05 (4 and 5 min before the prompt). When the step counts for different time lags were entered together as predictors in the same model, only the effect of step counts during the minute of the prompt remained significant (Table 3).

In supplemental analyses, we also examined whether physical activity levels predict the likelihood of missing (i.e., skipped) EMA responses. As EMA prompts expired after 15 min, response delays exceeding 15 min were treated as missed prompts. Paralleling the results for delayed responses, the probability of missing an EMA response significantly decreased to the extent that a respondent was more sedentary than usual, and missed EMA responses became significantly more likely when respondents spent more time in LPA, MPA, and VPA (see Supplementary Table 1).

Alternatively, it is also possible to conceptualize response delays as a time-to-event (i.e., duration) variable, that is, as the time lag from the EMA prompt to when the response occurred (see Elmer et al., in press). Results from time-to-event analyses using Cox proportional hazards models, which examined whether physical activity levels predicted the time to an EMA response, are presented in Supplementary Table 2. Missing responses were right censored, and random effects (frailty terms) were included to account for the clustered nature of the data. Results from the time-to-event analyses were consistent with the primary models, showing that sedentary behaviors predicted a higher probability (i.e., hazard) of responding to EMA prompts earlier, and spending more time in LPA, MPA, and VPA each predicted a lower probability of responding to EMA prompts earlier.

## Is delayed responding associated with changes in objective activity levels?

Results from multilevel models to estimate changes in respondents’ physical activity levels from the minute of an EMA prompt to the minute of a delayed response are shown in Table 4. There was no significant change in the amount of time standing

**TABLE 3** Results from multilevel logistic regression models predicting the occurrence of delayed responses from step counts recorded at increasing time lags before the EMA prompt.

Predictors	Odds ratio (95% CI)	
	Predictors entered individually in separate models	Predictors entered together in the same model
Step count during the minute of the EMA prompt	1.14 (1.10, 1.18)	1.12 (1.07, 1.17)
Step count 1 min before the EMA prompt	1.10 (1.07, 1.13)	1.02 (0.97, 1.07)
Step count 2 min before the EMA prompt	1.09 (1.05, 1.12)	1.02 (0.96, 1.07)
Step count 3 min before the EMA prompt	1.07 (1.04, 1.11)	1.02 (0.97, 1.08)
Step count 4 min before the EMA prompt	1.05 (1.02, 1.08)	0.97 (0.92, 1.03)
Step count 5 min before the EMA prompt	1.05 (1.01, 1.08)	1.00 (0.95, 1.05)

Odds ratios for which the 95% CI does not include 1.0 are significant at  $p < 0.05$ . CI, confidence interval.

from the time of the prompt to the time of the response. However, participants' time in sedentary states significantly increased from the time of the EMA prompt to the time of a delayed response. Moreover, participants exhibited significantly lower time in LPA, in MPA, in VPA, as well as a lower number of steps, comparing the minute of the prompt to the minute of the response during a response delay. The random coefficients for change in [Table 4](#) indicate substantial individual differences in the magnitude of these changes between individuals, with CVs ranging from 0.99 for step counts to 3.90 for time in sedentary states (except for VPA, CV = 0.26).

To put the magnitude of the changes in physical activity levels into context, in secondary analyses, we also explored how these changes compared to respondents' activity levels for prompts that were *not* delayed. To do this, we expanded the multilevel models for change by adding a categorical indicator representing immediate EMA responses as predictor variable. [Figure 3](#) shows the mean levels for each physical activity variable (a) at the time of the prompt (for delayed responses), (b) at the time of the response (for delayed responses), and (c) for immediate (i.e., nondelayed) responses to EMA prompts. As can be seen, compared to immediate responses, participants spent significantly less time in sedentary behaviors and more time standing at the time of a delayed response. However, for LPA, MPA, VPA, and step counts, the reduction in participants' activity levels during a response delay occurred to a level that was not significantly different on average from their activity levels for prompts that were answered immediately.

In supplemental moderator analyses, we also explored whether the magnitude of changes in activity levels from the time of the prompt to the time of a delayed response differed by the length of a delay (i.e., by the specific time difference between prompt and response). The specific delay duration did not significantly moderate the changes for any of the activity indicators (see [Supplementary Table 3](#)).

## Are response delays associated with systematic biases in self-reported activity levels?

Whether delayed responding translated into biases in EMA self-reports was examined next by comparing self-reported and objective physical activity levels across delayed and immediate responses. We first present the results for the primary analyses, that is, for respondents who were asked to rate their activities for the 5 min before the prompt. As shown in [Table 5](#), the estimated mean activity levels differed significantly between EMA reported and objective assessment types controlling for response delays; participants overall self-reported less sedentary time and more time standing and in LPA, MPA, and VPA, compared to objective recordings of these activities over the same time period. However, the assessment type by response delay interaction term was not significant for the variables examining sedentary time, standing, LPA, and VPA, indicating that the mean difference between EMA reported and objective activities did not significantly differ between immediate and delayed responses for these variables. Notably, there was substantial between-person variability in the magnitude and direction of the interaction terms. For MPA, the interaction term was significant, but the direction of the effect was contrary to the hypothesized effect: self-report ratings of MPA exceeded the objectively recorded MPA more when EMA responses were delayed compared to immediate responses, contrary to our hypothesis.

Similar results were found in secondary analyses of the group of respondents who completed EMA items asking about their activities over the 2 h before the prompt (see [Supplementary Table 4](#)). Participants self-reported less sedentary time and more time standing, in LPA, MPA, and VPA, compared to objective recordings of these activities over the 2-h period before the prompt. However, the assessment type by response delay interaction term was not significant for any of the physical activity variables, indicating that the differences in the levels of self-reported and objectively recorded activities did not differ according to whether a prompt was immediate or delayed.

## Are response delays associated with more random errors in self-reported activities?

Finally, we examined whether delayed responding moderated the magnitude of associations between self-reported and objective physical activities. We start with results for respondents completing the 5-min version of the EMA items. As shown in [Table 6](#), the results were similar for all physical activity variables except for VPA. For each variable, the regression coefficients of within-subject centered EMA reports predicting objective measurements of corresponding physical activities were highly significant and positive, indicating that self-reported and objective physical activities were positively associated. However, with the exception of VPA, this association was significantly moderated by response delays; the negative interaction term indicated that the association between EMA reported and objective activities was lower for delayed compared to immediate responses. The same pattern of effects was evident for EMA items presented with an open-ended

TABLE 4 Change in objective physical activity levels from the minute of the prompt to the minute of the response for delayed EMA responses.

Parameter	Sedentary (seconds)	Standing (seconds)	LPA (seconds)	MPA (seconds)	VPA (seconds)	Steps (count)
<b>Fixed effects</b>						
Intercept (time of prompt)	32.20 (0.92)***	18.94 (0.76)***	7.09 (0.43)***	1.41 (0.22)***	0.26 (0.11)*	11.76 (0.71)**
Change from prompt to response	2.37 (1.01)*	0.56 (0.90)	-2.28 (0.48)***	-0.44 (0.22)*	-0.24(0.11)*	-4.21 (0.77)***
<b>Variance components</b>						
Level 2 random effects						
Intercept	59.95	22.30	4.64	1.60	0.05	26.04
Change	85.36	49.69	7.74	0.65	0.00	17.56
Covariance	25.49	8.49	2.92	0.64	0.02	13.87
Level 1 residuals						
Time of prompt	677.40	527.44	85.24	27.51	0.02	219.25
Time or response	744.11	522.45	180.87	44.14	13.41	502.63
Covariance	-306.95	-197.86	-25.34	-6.85	-0.03	-55.24
<b>Fit statistics</b>						
-2 log likelihood	20,998.0	20,329.7	17,187.7	14,324.0	14,506.7	19,409.5
AIC	21,010.0	20,341.7	17,199.7	14,336.0	14,516.7	19,421.5
BIC	21,032.0	20,363.7	17,221.7	14,358.0	14,535.0	19,443.5

Standard errors are in parentheses. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ . LPA, light physical activity; MPA, moderate physical activity; VPA, vigorous physical activity; AIC, Akaike information criterion; BIC, Bayesian information criterion.

numeric response format and for EMA items using a “traditional” verbal rating scale.

To illustrate the magnitude of the moderated effects, Table 7 shows the within-person correlations between EMA reported and objective activities for each variable, separately for immediate and delayed responses. The correlations ranged between  $r = 0.04$  (for time in VPA) to  $r = 0.70$  (time sedentary) for immediate responses and between  $r = 0.00$  (time in MPA) to  $r = 0.65$  (time sedentary) for delayed responses. Notably, delayed reports of MPA showed virtually no correlation with participants’ objective MPA at the time of the prompt, suggesting that these reports were no more informative about objective MPA than random responses. Effect sizes for differences in the Fisher  $z$ -transformed correlations (i.e., Cohen’s  $q$ ) between immediate versus delayed responses are also shown in Table 7, where values of  $q = 0.10$ ,  $0.30$ , and  $0.50$  suggest small, medium, and large effects, respectively (Cohen, 1988). Effect sizes ranged between  $q = 0.08$  to  $q = 0.22$  (except for VPA;  $q = 0.06$ ), suggesting overall small effects. The negative effect size for VPA indicates that, contrary to expectation, delayed responses were somewhat more highly correlated with actual behaviors compared to immediate responses, though this difference was not statistically significant.

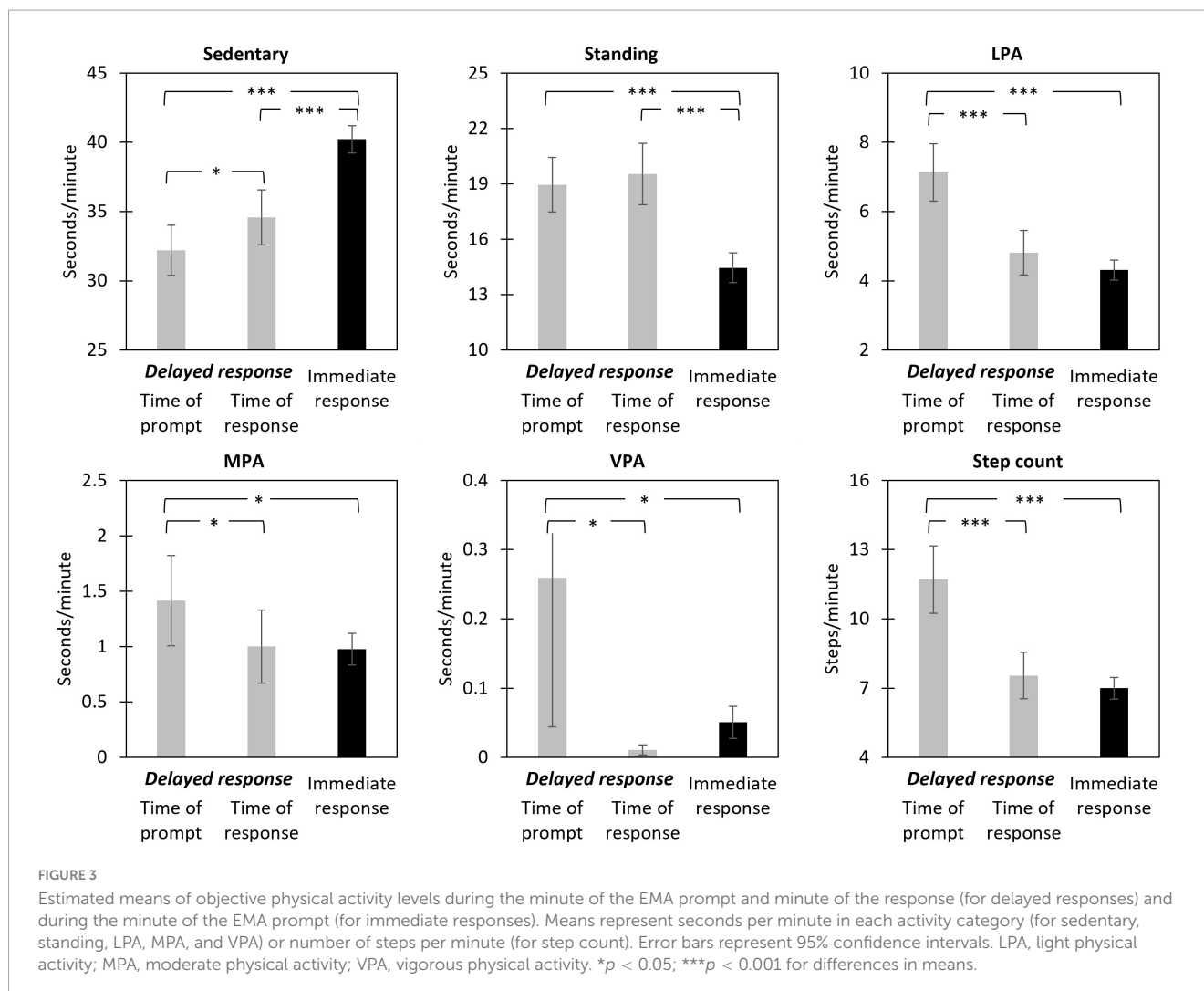
In secondary analyses of respondents who reported their activities over the 2 h before the prompt, response delays did not significantly moderate the association between self-reported and objective activities for most variables except for LPA (see Supplementary Table 5). For time spent in LPA, the within-person correlation between self-reported and objectively recorded activities was  $r = 0.38$  for immediate and  $r = 0.27$  for delayed responses ( $q = 0.12$  for the difference in these correlations, see Supplementary Table 6).

## Discussion

A major advantage of EMA is the ability to collect information about experiences and behaviors in (near) real time. For the methodology to provide optimal results, participants must provide timely EMA reports when prompted. Delayed EMA responses create a temporal misalignment between the occurrence of a behavior and its reporting. However, to date, the conditions generating the misalignment and the consequences of the temporal misalignment for the quality of EMA reports are poorly understood. Using continuous activity monitoring as an objective reference value, we found that participants systematically self-selected the moments when they answered EMA surveys based on their current physical activity behaviors. There was only very limited evidence to suggest that this self-selection of the moments of reporting translated into biases in EMA reports, although random error was increased in delayed reports.

We found that participants on average delayed about 10% of EMA reports by more than 1 min. The likelihood of delays was lower when individuals spent more time in sedentary states and increased to rates of 20% or more for higher activity levels. This finding extends prior literature documenting that contextual and behavioral factors beyond structural barriers to EMA reporting (such as driving, sleeping) can contribute to EMA response delays (Sarker et al., 2014; Boukhechba et al., 2018), and that participants are more likely to skip prompts entirely when they are more physically active (Dunton et al., 2012). It is noteworthy that even small elevations of activity levels (time standing, time in LPA) above sedentary behaviors predicted a greater chance of response delays, suggesting that delays were not exclusively attributable to vigorous physical exercise (e.g., jogging, running).





Importantly, both within-person fluctuations and between-person differences in objective physical activity levels contributed in a similar way to the occurrence of delayed responding. It has often been pointed out that failing to distinguish within- and between-person effects can result in “uninterpretable blends” between two possibly very different underlying mechanisms (e.g., Hamaker, 2012). Our results suggest that physical activity monitoring with wearable devices is comparably relevant for understanding and predicting participants who will more likely engage in delayed responding and for predicting the moments and EMA prompts for which a given person will more likely delay a response.

Although the majority of EMA research relies on stratified random prompts to collect self-report data that is most representative of all waking hours of a participant’s day, event-contingent EMA prompts have been used to gain insights into people’s experiences in specific contexts. Increasingly, researchers have leveraged passive activity sensors to trigger event-contingent EMA prompts in “context-aware” sampling designs (Aminikhanghahi et al., 2019; Hoemann et al., 2020). For example, Dunton et al. (2014) proposed using motion sensors integrated in smartphones to detect likely bouts of sedentary and physically active behaviors, and to use times between these

bouts to trigger EMA surveys to gather specific information about where, with whom, and why these physical activity states occurred. Context-aware delivery of EMA prompts triggered by activity sensors has also been explored in attempts to limit participant burden, increase response rates, and facilitate timely EMA responses (Khanshan et al., 2021; King et al., 2024). Our results indicated that higher activity levels (greater step counts) monotonically increased the chances of response delays, but they did not reveal a threshold of activity levels that would yield categorically different response behaviors. Moreover, even though lagged effects analyses provided some evidence that response delays were associated with activity behaviors occurring before an EMA prompt, the effect sizes of lagged effects of these behaviors dwindled quickly within just a few minutes before a prompt. This suggests that attempts to predict when respondents may be most responsive to EMA prompts based on passive activity sensors may be most successful when considering gradual and transient changes in physical activity levels.

As hypothesized, we found that respondents’ activity levels changed during response delays. Specifically, participants appeared to transition from either light, moderate, or vigorous activities to less active (standing or sedentary) states that were similar to

TABLE 5 Results of multilevel models comparing objective and EMA reported physical activity levels for the 5 min before the prompt across immediate and delayed responses.

Parameter (scaled in seconds per minute)	Sedentary	Standing	LPA	MPA	VPA
<b>Fixed effects</b>					
Intercept	38.54 (0.58)***	15.25 (0.49)***	11.83 (0.31)***	1.79 (0.12)***	.65 (0.10)***
Assessment type	−4.13 (0.49)***	2.51 (0.63)***	14.41 (0.52)***	1.73 (0.23)***	1.21 (0.21)***
Response delay	−5.18 (1.07)***	3.40 (0.84)***	1.75 (0.64)**	1.26 (0.52)*	.41 (0.19)*
Assessment type × delay	−0.37 (1.01)	0.51 (1.00)	0.63 (1.04)	0.51 (0.30)*	0.55 (0.36)
<b>Variance components</b>					
Level 2 random effects					
Intercept	41.58	29.77	10.25	1.21	1.38
Assessment type	27.75	56.30	29.41	5.47	5.68
Response delay	4.36	1.10	1.24	0.00 <sup>a</sup>	0.42
Assessment type × delay	17.49	15.39	0.11	0.00 <sup>a</sup>	0.72
<b>Level 1 residuals</b>					
Accelerometry assessments	565.25	362.83	62.50	25.81	2.14
EMA assessments	629.35	470.34	543.66	118.82	55.15
Covariance	415.97	231.20	58.81	10.20	1.00
<b>Fit statistics</b>					
−2 log likelihood	90,814.5	88,916.5	82,290.3	70,469.0	54,829.8
AIC	90,828.5	88,928.5	82,302.3	70,479.0	54,843.8
BIC	90,850.6	88,947.4	82,321.2	70,494.7	54,865.9

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ . Assessment type = EMA reports versus objective recordings. Response delay = delayed versus immediate EMA responses. LPA, light physical activity; MPA, moderate physical activity; VPA, vigorous physical activity. <sup>a</sup>Variance components were fixed at zero to achieve a positive definite covariance matrix of the random effects.

activity levels for prompts that were *not* delayed. The duration of the response delay did not significantly moderate the amount of change in activities, suggesting that respondents may have tended to postpone responding until they were in a less active state, regardless of how long it took them to reduce their activity levels.

However, analyses of participants' EMA ratings did not show that this change in objective activity levels systematically affected self-reported activity levels either for participants who were asked about the 5 min or about the 2 h before the prompt. In line with prior research (e.g., Lim et al., 2015; Prince et al., 2020; Quinlan et al., 2021), participants tended to generally underreport sedentary time and to overreport the amount of physically active time in EMA compared to objective measurements of the same behaviors before the prompts. These differences were not, however, moderated by response delays, such that the EMA responses did not differ from what would be expected had the responses not been delayed. That is, delayed responses did not lead to a selective representation of moments of high (or low) physical activity levels in the collected EMA data.

It is noteworthy that the present study explicitly asked respondents to recall and report their activity levels during the minutes "before the prompt." An alternative instruction frequently used in EMA studies is to ask participants to report their "current" experiences and behaviors (Hall et al., 2021). Even though the latter instruction arguably minimizes potential recall biases compared with the former (Singh and Björling, 2019), it might have yielded a more pronounced impact of response delays on the

self-reported activity levels as respondents' current objective state (i.e., activity level) differed from that at the time of the prompt. However, this possibility remains speculative and will likely depend on how respondents interpret specific EMA instructions. For example, a previous study employed brief telephone-based cognitive interviews several times a day at random intervals to investigate which time periods EMA respondents would take into account for their momentary ratings (Wen et al., 2021). Results showed that unless participants were specifically trained to focus on the time right before the telephone call, only about half of the participants closely adhered to the time period as intended by the instructions. The extent to which EMA instructions modulate the impact of delayed responding will need to be empirically tested in future (experimental and qualitative interview) studies.

Examining the correlations between objective and EMA reported physical activities, we found evidence to suggest that delayed EMA reporting was associated with more random errors in measurement when respondents were asked about the 5 min before the prompt. Immediate EMA reports showed moderate correlations between objective and EMA reported activity data, corresponding with prior research (Knell et al., 2017; Stinson et al., 2022), whereas these correlations were significantly attenuated for delayed EMA reports. That said, the effect sizes for the differences in correlations were generally small, suggesting that potential measurement errors resulting from delayed responses are not likely to markedly reduce the quality and precision of the overall EMA data collected, especially when the rate of

TABLE 6 Multilevel regression models predicting objective physical activity levels during the 5 min before the EMA prompt from EMA reported physical activity levels, for immediate versus delayed EMA responses.

Parameter (scaled in seconds per minute)	Physical activity outcomes for EMA items with open numeric response format					Physical activity outcomes for EMA items with rating scale format <sup>a</sup>	
	Sedentary	Standing	LPA	MPA	VPA <sup>b</sup>	Sedentary	LMVPA
<b>Fixed effects</b>							
Intercept	40.34 (0.67)***	13.82 (0.52)***	4.58 (0.19)***	0.94 (0.09)***	0.50 (0.32)	40.87 (0.66)***	5.41 (0.22)***
Delay	-1.11 (0.82)	1.99 (0.76)**	1.12 (0.36)**	0.07 (0.22)	0.03 (0.05)	-1.44 (0.92)	0.03 (0.39)
EMA	0.65 (0.02)***	0.50 (0.02)***	0.12 (0.01)***	0.08 (0.02)***	0.06 (0.05)	10.55 (0.36)***	5.44 (0.26)***
Delay × EMA	-0.07 (0.03)*	-0.11 (0.04)**	-0.04 (0.02)*	-0.08 (0.02)***	0.01 (0.01)	-1.27 (0.65)*	0.90 (0.40)*
<b>Random effects</b>							
Intercept	67.10	37.73	4.11	0.76	17.90	62.54	6.30
EMA regression slope	0.06	0.04	0.01	0.04	0.36	15.18	8.39
Covariance	-0.69	0.38	0.05	0.15	2.53	-12.62	5.56
Level 1 residual	255.73	229.27	51.30	20.12	0.77	323.43	58.96
<b>Fit statistics</b>							
-2 log likelihood	40,786.3	40,304.7	33,065.7	28,746.9	13,696.0	42,226.1	34,377.2
AIC	40,794.3	40,312.7	33,073.7	28,754.9	13,704.0	42,234.1	34,385.2
BIC	40,806.9	40,325.3	33,086.3	28,767.5	13,716.6	42,246.7	34,397.9

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ . LPA, light physical activity; MPA, moderate physical activity; VPA, vigorous physical activity; LMVPA, light, moderate, or vigorous physical activity. <sup>a</sup>EMA rating scale items were “during the 5 min before the prompt, how sedentary were you?” and “during the 5 min before the prompt, how physically active were you?” <sup>b</sup>Accelerometry and EMA data for VPA were log transformed to facilitate model convergence.

TABLE 7 Within-subject correlations between objective physical activity levels during the 5 min before the EMA prompt and EMA reported physical activity levels, for immediate and delayed EMA responses.

Physical activity variable	Correlation between objective and EMA self-reported activities		Difference in correlations
	Immediate response	Delayed response	Effect size $q$
Time sedentary	0.70	0.65	0.09
Time standing	0.57	0.46	0.15
Time in LPA	0.33	0.23	0.11
Time in MPA	0.21	0.00	0.21
Time in VPA	0.04	0.10	-0.06
Rating of how sedentary	0.65	0.57	0.13
Rating of how physically active	0.54	0.49	0.07

delayed responses is modest. However, this finding is consistent with the idea that memories of past events and behaviors fade quickly over time and that imprecisions in the recall of behaviors occur after very few minutes of delay time (Conner and Barrett, 2012). It is important to keep in mind that the maximum allowable delay in the present study was 15 min, and that EMA studies vary widely in the maximum response delay and have allowed response latencies of 30 min or longer (Eisele et al., 2021). Even though the present results cannot confirm this, it is possible that response delays of more than

15 min would be associated with even more pronounced reporting errors.

In secondary analyses conducted with EMA items that asked about activity levels over 2 h before the prompt rather than examining the shorter 5-min reporting period, we found very little evidence that delayed EMA responding was associated with attenuated correlations between objective and self-reported activities. Possible advantages and disadvantages of using EMA items with longer reporting periods such as the “past 2 h” (the so-called “coverage” model of EMA) have been discussed in

Stone et al. (2023a). As suggested by the present results, a notable feature of longer reporting periods is that they are likely robust to potential reporting errors when individuals need to delay prompt responses, especially if the length of delays is brief (e.g., up to 15 min) relative to the time period covered by the EMA report (e.g., past 2 h). However, we acknowledge that this advantage may be offset if participants struggle to remember the entire time period covered by the EMA question, such that memory heuristics (e.g., peak and end heuristics) that are unrelated to the occurrence of response delays may come into play.

When deciding on the maximum allowable response delay in an EMA study, researchers may ask whether setting delayed responses to missing values would help reduce potentially biased estimates. Our results suggest that allowing for response delays of 15 min or less may not substantially bias EMA data on self-reported physical activity. At the same time, because respondents systematically delayed responses when they were more active, *not* collecting delayed responses (or discarding them from the data) could come at the cost of introducing missing values that are not missing (completely) at random, which could bias study results (Stone et al., 2023a).

## Limitations

The study sample was drawn from a larger general US population sample of internet panelists who regularly participated in online surveys. Prior research has shown that the samples recruited from these panels are not necessarily representative of the general population (Stone et al., 2023b) and the observed effects may be different in other (clinical) populations which may differ in the patterns of delayed EMA responses and in their motivation for timely assessment completion. As EMA is increasingly applied in clinical populations, further research on the occurrence and impact of delayed responses in clinical samples is needed.

Our study exclusively focused on physical activity behaviors, and the results may not necessarily translate to EMA reports of other behaviors or internal states. Focusing on physical activities provided a near-unique opportunity to compare EMA reports with continuous real-time recordings of the same (or closely related) concept. With few notable exceptions, constructs typically evaluated with EMA lack objective equivalents. For instance, health behaviors such as smoking and alcohol use that can be indirectly monitored in ambulatory settings (McClure et al., 2018), and environmental features such as noise levels can be passively recorded in real-time (Timmer et al., 2017). Moreover, biological correlates of psychological states such as stress have been inferred from ambulatory physiological data (e.g., skin conductance, heart rate variability), even though a close approximation of internal states remains challenging (Smets et al., 2018).

This study is limited by the use of the activPAL device, which primarily captures lower-body movements. Moderate or vigorous physical activities involving upper-body movements or minimal lower-body engagement, such as weightlifting or rowing, may have been missed as they are not fully captured by this thigh-worn device. In addition, the resolution of accelerometry-derived physical activities was 1 min, which may have introduced error variance due to some uncertainty about activity levels at the exact time of an EMA prompt.

The present study was also limited to two specific EMA designs that differed in the time period targeted in the EMA reports. Multiple characteristics of the prompting schedule (prompt frequency, number and timing of follow-up alarms, maximum allowable delay time) may act in concert to shape the occurrence and duration of delayed EMA responses, and differences in the design of EMA questions (time period covered by EMA questions, instructions about reporting the moments before the prompt or the current moments) may affect how response delays impact the precision and accuracy of EMA reports. Future studies are needed to investigate the occurrence and impact of response delays for various study design features.

## Conclusion

Delayed responses are a frequently overlooked “subtle” form of noncompliance in EMA studies. Participants in this study were more likely to delay EMA responses when they were more physically active, and responded after activity levels were significantly lower. Brief response delays did not systematically bias the physical activity levels captured with EMA, although they were associated with more random errors in EMA reports, with generally small effect sizes. To empirically inform best practices regarding the treatment of delayed responses in EMA studies, further observational and experimental research should investigate the generalizability of effects of delayed EMA responses across populations and EMA contents. Investigations in clinical populations are warranted considering that patients may differ from nonpatients in the frequency and reasons for response delays (such as physical symptoms that impede timely assessment completion). In addition, examining the effects of delayed responses across EMA contents may be especially important considering that concrete behaviors may be more memorable and less impacted by response delays compared to internal states that may be most accessible in the moments they are experienced. Finally, studies should systematically compare different features of EMA study designs, for example, to examine whether certain instructions for EMA reporting (“right now” versus “before the prompt”) are more susceptible to biases from delayed responses, and to investigate the quality of EMA responses that are substantially delayed (e.g., by more than 15 min).

## Data availability statement

The data are not publicly available, but can be shared upon request. Requests to access the data should be directed to AS, [arthuras@usc.edu](mailto:arthuras@usc.edu).

## Ethics statement

The studies involving humans were approved by the Biomedical Research Alliance of New York Institutional Review Board. The studies were conducted in accordance with the



local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

SS: Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. MT: Conceptualization, Data curation, Methodology, Project administration, Supervision, Validation, Writing – review & editing. DJ: Conceptualization, Project administration, Supervision, Writing – review & editing. JS: Conceptualization, Methodology, Writing – review & editing. P-JL: Data curation, Writing – review & editing. SG: Data curation, Project administration, Supervision, Validation, Writing – review & editing. OP: Data curation, Project administration, Writing – review & editing. AS: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by a grant from the National Institute on Aging, R37AG057685.

## References

- Affleck, G., Tennen, H., Urrows, S., Higgins, P., Abeles, M., Hall, C., et al. (1998). Fibromyalgia and women's pursuit of personal goals: a daily process analysis. *Health Psychol.* 17, 40–47. doi: 10.1037//0278-6133.17.1.40
- Aiken, L. S., West, S. G., and Reno, R. R. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, CA: Sage.
- Alattar, L., Messel, M., and Rogofsky, D. (2018). An introduction to the understanding America Study internet panel. *Soc. Secur. Bull.* 78, 13–28.
- Aminikhanghahi, S., Schmitter-Edgecombe, M., and Cook, D. J. (2019). Context-aware delivery of ecological momentary assessment. *IEEE J. Biomed. Health Inform.* 24, 1206–1214.
- Bassett, D. R., John, D., Conger, S. A., Rider, B. C., Passmore, R. M., and Clark, J. M. (2014). Detection of lying down, sitting, standing, and stepping using two activPAL monitors. *Med. Sci. Sports Exerc.* 46, 2025–2029. doi: 10.1249/MSS.0000000000000326
- Bekman, N. M., Winward, J. L., Lau, L. L., Wagner, C. C., and Brown, S. A. (2013). The impact of adolescent binge drinking and sustained abstinence on affective state. *Alcohol. Clin. Exp. Res.* 37, 1432–1439. doi: 10.1111/acer.12096
- Bolger, N., and Laurenceau, J.-P. (2013). *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*. New York, NY: Guilford Press.
- Boukhechba, M., Cai, L., Chow, P. I., Fua, K., Gerber, M. S., Teachman, B. A., et al. (2018). "Contextual analysis to understand compliance with smartphone-based ecological momentary assessment," in *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, New York, NY, 232–238.
- Businelle, M. S., Ma, P., Kendzor, D. E., Frank, S. G., Vidrine, D. J., and Wetter, D. W. (2016). An ecological momentary intervention for smoking cessation: evaluation of feasibility and effectiveness. *J. Med. Internet Res.* 18:e321.
- Christensen, T. C., Barrett, L. F., Bliss-Moreau, E., Lebo, K., and Kaschub, C. (2003). A practical guide to experience-sampling procedures. *J. Happ. Stud.* 4, 53–78.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Conner, T. S., and Barrett, L. F. (2012). Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. *Psychosom. Med.* 74, 327–337.
- Courtney, J., Nuss, K., Lyden, K., Harrall, K., Glueck, D., Villalobos, A., et al. (2021). Comparing the activPAL software's primary time in bed algorithm against self-report and van der berg's algorithm. *Meas. Phys. Educ. Exerc. Sci.* 25, 212–226. doi: 10.1080/1091367x.2020.1867146
- De Vries, L. P., Baselmans, B. M., and Bartels, M. (2021). Smartphone-based ecological momentary assessment of well-being: a systematic review and recommendations for future studies. *J. Happ. Stud.* 22, 2361–2408. doi: 10.1007/s10902-020-00324-7
- Do, B., Mason, T. B., Yi, L., Yang, C.-H., and Dunton, G. F. (2021). Momentary associations between stress and physical activity among children using ecological momentary assessment. *Psychol. Sport Exerc.* 55:101935.
- Dunton, G. F. (2017). Ecological momentary assessment in physical activity research. *Exerc. Sport Sci Rev.* 45, 48–54.
- Dunton, G. F., Dzubur, E., Kawabata, K., Yanez, B., Bo, B., and Intille, S. (2014). Development of a smartphone application to measure physical activity using sensor-assisted self-report. *Front. Public Health* 2:12. doi: 10.3389/fpubh.2014.00012
- Dunton, G. F., Liao, Y., Kawabata, K., and Intille, S. (2012). Momentary assessment of adults' physical activity and sedentary behavior: feasibility and validity. *Front. Psychol.* 3:260. doi: 10.3389/fpsyg.2012.00260

## Conflict of interest

AS is a senior scientist with the Gallup Organization and a consultant for HRA Pharma.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1503411/full#supplementary-material>

- Eisele, G., Vachon, H., Myin-Germeys, I., and Viechtbauer, W. (2021). Reported affect changes as a function of response delay: findings from a pooled dataset of nine experience sampling studies. *Front. Psychol.* 12:580684. doi: 10.3389/fpsyg.2021.580684
- Elmer, T., Van Duijn, M. A., Ram, N., and Bringmann, L. F. (in press). Modeling categorical time-to-event data: the example of social interaction dynamics captured with event-contingent experience sampling methods. *Psychol. Methods*. [Epub ahead of print]. doi: 10.1037/met0000598
- Graham, J. (2012). *Missing Data: Analysis and Design*. New York, NY: Springer.
- Hall, M., Scherner, P. V., Kreidel, Y., and Rubel, J. A. (2021). A systematic review of momentary assessment designs for mood and anxiety symptoms. *Front. Psychol.* 12:642044. doi: 10.3389/fpsyg.2021.642044
- Hamaker, E. L. (2012). "Why researchers should think "within-person": a paradigmatic rationale," in *Handbook of Research Methods for Studying Daily Life*, eds M. R. Mehl and T. S. Conner (New York, NY: Guilford), 43–61.
- Hoemann, K., Khan, Z., Feldman, M. J., Nielson, C., Devlin, M., Dy, J., et al. (2020). Context-aware experience sampling reveals the scale of variation in affective experience. *Sci. Rep.* 10:12459.
- Jones, A., Remmerswaal, D., Verveer, I., Robinson, E., Franken, I. H., Wen, C. K. F., et al. (2019). Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. *Addiction* 114, 609–619. doi: 10.1111/add.14503
- Khanshan, A., Van Gorp, P., Nuijten, R., and Markopoulos, P. (2021). Assessing the influence of physical activity upon the experience sampling response rate on wrist-worn devices. *Int. J. Environ. Res. Public Health* 18:10593. doi: 10.3390/ijerph182010593
- King, Z. D., Yu, H., Vaessen, T., Myin-Germeys, I., and Sano, A. (2024). Investigating receptivity and affect using machine learning: ecological momentary assessment and wearable sensing study. *JMIR mHealth uHealth* 12:e46347. doi: 10.2196/46347
- Knell, G., Gabriel, K. P., Businelle, M. S., Shuval, K., Wetter, D. W., and Kendzor, D. E. (2017). Ecological momentary assessment of physical activity: validation study. *J. Med. Internet Res.* 19:e253.
- Lim, S., Wyker, B., Bartley, K., and Eisenhower, D. (2015). Measurement error of self-reported physical activity levels in New York City: assessment and correction. *Am. J. Epidemiol.* 181, 648–655. doi: 10.1093/aje/kwv470
- Lyden, K., Keadle, S. K., Staudenmayer, J., and Freedson, P. S. (2017). The activPAL™ accurately classifies activity intensity categories in healthy adults. *Med. Sci. Sports Exerc.* 49, 1022–1028. doi: 10.1249/MSS.0000000000001177
- McClure, E. A., Tomko, R. L., Carpenter, M. J., Treiber, F. A., and Gray, K. M. (2018). Acceptability and compliance with a remote monitoring system to track smoking and abstinence among young smokers. *Am. J. Drug Alcohol Abuse* 44, 561–570. doi: 10.1080/00952990.2018.1467431
- Muthén, L. K., and Muthén, B. O. (2017). *Mplus: Statistical Analysis with Latent Variables: User's Guide (Version 8)*. Los Angeles, CA: Muthén & Muthén.
- Ono, M., Schneider, S., Junghaenel, D. U., and Stone, A. A. (2019). What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. *J. Med. Internet Res.* 21:e11398. doi: 10.2196/11398
- Podsakoff, N. P., Spoelma, T. M., Chawla, N., and Gabriel, A. S. (2019). What predicts within-person variance in applied psychology constructs? An empirical examination. *J. Appl. Psychol.* 104, 727–754. doi: 10.1037/apl0000374
- Prince, S. A., Cardilli, L., Reed, J. L., Saunders, T. J., Kite, C., Douillette, K., et al. (2020). A comparison of self-reported and device measured sedentary behaviour in adults: a systematic review and meta-analysis. *Int. J. Behav. Nutr. Phys. Activ.* 17:31.
- Quinlan, C., Rattray, B., Pryor, D., Northey, J. M., Anstey, K. J., Butterworth, P., et al. (2021). The accuracy of self-reported physical activity questionnaires varies with sex and body mass index. *PLoS One* 16:e0256008. doi: 10.1371/journal.pone.0256008
- Sarker, H., Sharmin, M., Ali, A. A., Rahman, M. M., Bari, R., Hossain, S. M., et al. (2014). "Assessing the availability of users to engage in just-in-time intervention in the natural environment," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Seattle, WA, 909–920. doi: 10.1145/2632048.2636082
- Schneider, S., and Stone, A. A. (2016). Ambulatory and diary methods can facilitate the measurement of patient-reported outcomes. *Qual. Life Res.* 25, 497–506.
- Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4, 1–32.
- Singh, N. B., and Björling, E. A. (2019). A review of EMA assessment period reporting for mood variables in substance use research: expanding existing EMA guidelines. *Addict. Behav.* 94, 133–146. doi: 10.1016/j.addbeh.2019.01.033
- Smets, E., De Raedt, W., and Van Hoof, C. (2018). Into the wild: the challenges of physiological stress detection in laboratory and ambulatory settings. *IEEE J. Biomed. Health Inform.* 23, 463–473. doi: 10.1109/JBHI.2018.2883751
- Snijders, T. B., and Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.
- Stinson, L., Liu, Y., and Dallery, J. (2022). Ecological momentary assessment: a systematic review of validity research. *Perspect. Behav. Sci.* 45, 469–493.
- Stone, A. A., and Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Ann. Behav. Med.* 16, 199–202.
- Stone, A. A., Schneider, S., and Smyth, J. M. (2023a). Evaluation of Pressing issues in ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 19, 107–131. doi: 10.1146/annurev-clinpsy-080921-083128
- Stone, A. A., Schneider, S., Smyth, J. M., Junghaenel, D. U., Wen, C., Couper, M. P., et al. (2023b). Shedding light on participant selection bias in Ecological Momentary Assessment (EMA) studies: findings from an internet panel study. *PLoS One* 18:e0282591. doi: 10.1371/journal.pone.0282591
- Thurman, S. M., Wasylshyn, N., Roy, H., Lieberman, G., Garcia, J. O., Asturias, A., et al. (2018). Individual differences in compliance and agreement for sleep logs and wrist actigraphy: a longitudinal study of naturalistic sleep in healthy adults. *PLoS One* 13:e0191883. doi: 10.1371/journal.pone.0191883
- Timmer, B. H., Hickson, L., and Launer, S. (2017). Ecological momentary assessment: feasibility, construct validity, and future applications. *Am. J. Audiol.* 26, 436–442.
- Tudor-Locke, C., and Rowe, D. A. (2012). Using cadence to study free-living ambulatory behaviour. *Sports Med.* 42, 381–398.
- Vachon, H., Viechtbauer, W., Rintala, A., and Myin-Germeys, I. (2019). Compliance and retention with the experience sampling method over the continuum of severe mental disorders: meta-analysis and recommendations. *J. Med. Internet Res.* 21:e14475. doi: 10.2196/14475
- Walentynowicz, M., Schneider, S., and Stone, A. A. (2018). The effects of time frames on self-report. *PLoS One* 13:e0201655. doi: 10.1371/journal.pone.0201655
- Wen, C. K. F., Junghaenel, D. U., Newman, D. B., Schneider, S., Mendez, M., Goldstein, S. E., et al. (2021). The effect of training on participant adherence with a reporting time frame for momentary subjective experiences in ecological momentary assessment: cognitive interview study. *JMIR Format. Res.* 5:e28007. doi: 10.2196/28007
- Wen, C. K. F., Liao, Y., Maher, J. P., Huh, J., Belcher, B. R., Dzubur, E., et al. (2018). Relationships among affective states, physical activity, and sedentary behavior in children: moderation by perceived stress. *Health Psychol.* 37, 904–914. doi: 10.1037/hea0000639
- Wrzus, C., and Neubauer, A. B. (2023). Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. *Assessment* 30, 825–846.
- Yaremych, H. E., Preacher, K. J., and Hedeker, D. (2023). Centering categorical predictors in multilevel models: best practices and interpretation. *Psychol. Methods* 28, 613–630. doi: 10.1037/met0000434
- Zlatar, Z. Z., Campbell, L. M., Tang, B., Gabin, S., Heaton, A., Higgins, M., et al. (2022). Daily level association of physical activity and performance on ecological momentary cognitive tests in free-living environments: a mobile health observational study. *JMIR mHealth uHealth* 10:e33747. doi: 10.2196/33747