Standard Length Article

Sociological Methodology
1–19

DOI: 10.1177/00811750261419100
http://sm.sagepub.com

Comparing Ecological Momentary Assessments and Time Diary Methods for Measuring Daily Life

Siyun Peng[1], Brea L. Perry[2] and Adam R. Roth[3]

Abstract

A growing number of social scientists are using ecological momentary assessment (EMA) to observe how social forces operate in real time. However, the validity of EMA for measuring features of daily life—what people are doing, where, and with whom—remains uncertain. A key challenge is the lack of consensus across studies about how validity in EMA methods is defined and assessed. The authors address that gap by comparing EMA data (n = 1,174) with time diary data (n = 1,113) using two population-based samples. An advantage of large samples is the ability to evaluate the magnitude of bias rather than relying solely on p-values, as is common in small-sample studies. The authors find that both methods yield similar estimates of moments captured at home and in the workplace, supporting their validity in those contexts. However, EMA tends to overestimate moments spent alone compared with time diaries, likely because of moment selection bias. Moreover, large discrepancies in estimates for eating and drinking and household chores suggest that relying on primary activity reports can introduce significant bias for multitasked activities. Comparing these methods provides insight into their relative strengths and limitations, helping researchers assess the validity, potential biases, and interpretive implications of each across key domains of daily life.

Keywords

ecological momentary assessments, time diary, daily life, moment selection bias, validity, bias

Understanding how individuals allocate their time in daily life is central to many sociological inquiries, including research on work and family dynamics, gender inequality, health, well-being, and the organization of social roles (Andorka 1987; Cornwell, Gershuny, and Sullivan 2019; Gerstel and Clawson 2018; Lundberg, Komarovsky, and McInerny 1934; Roth 2024). However, gathering self-reported data on daily life presents a number of methodological challenges. The most commonly used approach involves stylized survey questions where respondents (or proxies) are asked to estimate the "typical" or "average" amount of time they spend on specific activities over a given

[1]School of Aging Studies, University of South Florida, Tampa, Florida, USA
[2]Department of Sociology and Irsay Institute, Indiana University, Bloomington, Indiana, USA
[3]Department of Sociology, Oklahoma State University, Stillwater, Oklahoma, USA

Corresponding Author:
Siyun Peng, University of South Florida, School of Aging Studies, 13301 Bruce B. Downs Boulevard, MHC 1349, Tampa, FL 33612, USA
Email: siyunpeng@usf.edu

period, such as a day, week, month, or year. For example, a respondent might be asked, "On average over the past 12 months, how often have you participated in vigorous physical activity or exercise?" This method requires individuals to summarize and categorize their behavior over long periods of time, effectively aggregating many discrete events into generalized estimates. Such aggregation places a significant cognitive burden on respondents and increases the risk for measurement error, as people may misremember, misclassify, or misestimate the frequency or duration of their activities (Juster, Ono, and Stafford 2003).

A growing number of researchers are therefore turning to ecological momentary assessment (EMA) as a strategy to minimize the recall bias inherent in stylized questions (Browning et al. 2024; Roth 2024). EMA offers a powerful tool for observing how social forces operate in real time, a key attraction for sociological research. Yet the validity of EMA for measuring daily life—specifically, capturing what individuals are doing, where they are, and with whom they are interacting—remains an open question (Stinson, Liu, and Dallery 2022). In the present study we address this gap by comparing estimates between EMA and time diaries, the two most commonly used methods for capturing moments in daily life. This allows us to assess the validity, bias, and interpretive implications of each method. This work also provides insight into how sociological concepts—such as social connectedness, labor, and inequalities—manifest differently depending on the tools used to observe them.

MEASUREMENT VALIDITY AND BIAS IN EMA

EMA, also known as experience sampling, is a data collection method that asks respondents to report on their thoughts, feelings, behaviors, activities, or environments in (or near) real time, typically through randomized or structured prompts delivered throughout the day (Browning et al. 2024). In a typical EMA design, individuals are pinged every few hours to report on current experiences that would be difficult to recall accurately later. This methodology has provided many useful insights into a range of social processes, including emotional states, pain levels, and health behaviors (Monnaatsie et al. 2024; Overton et al. 2023; Solhan et al. 2009). However, a recent review of EMA's validity found that agreement rates between EMA reports and objective measures (e.g., wearable cameras, accelerometers, and direct observation) varied considerably across behaviors and studies, ranging from as low as 1.8 percent to as high as 100 percent (Stinson et al. 2022).

In their review, Stinson et al. (2022) highlighted the lack of consistency across studies regarding what constitutes validity. For example, some studies found relatively large differences between EMA and objective measures but nevertheless concluded that the EMA measures were valid because the corresponding p-values were not statistically significant. Other studies have reported similar differences but have instead concluded that EMA measures are invalid because of their statistically significant p-values. These conflicting conclusions demonstrate that relying on statistical significance as the standard for establishing validity is problematic (Sullivan and Feinn 2012), particularly given the small sample sizes typical of most validity studies (usually $n < 100$).

It is important to note that most validity studies of EMA, which come largely from behavioral medicine and psychology, focus on how accurately the method captures specific types of activities, such as physical activity, substance use, or dietary behaviors (Monnaatsie et al. 2024; Noh et al. 2025; Stinson et al. 2022). However, sociologists have begun using EMA to measure the full range of activities of daily life (e.g., what individuals are doing, where they are, and who they interact with). Examples include the 500 Family Study (Larson and Richards 1994), the Adolescent Health and Development in Context study (Browning et al. 2021), and the Chicago Health and Activity in Real Time study (Compernolle et al. 2022). Compared with traditional uses of EMA that target a specific behavior or experience, using EMA to capture all types of daily activities through randomized prompts is more complicated regarding potential biases. As a result, the validity of using EMA for measuring daily life remains an open question. In the current study, we aim to conduct a thoughtful examination of validity issues to enhance the overall utility of EMA, including improving the quality of inferences made using EMA data.

One important source of bias in EMA is missing responses or moment selection bias. A meta-analysis of 494 EMA samples reported an average compliance rate of 79 percent (i.e., the percentage of answered prompts out of all scheduled assessments), with a standard deviation of 13.64 percent (Wrzus and Neubauer 2023). Missing responses are common in EMA studies because respondents are often required to respond to multiple prompts per day while engaged in various activities. Missing responses do not produce biased estimates when they are missing at random. In this case, missing EMA data only lead to a loss of information and decreased statistical power. However, if missingness is not random—for instance, if prompts are more likely to be missed when respondents are busy, tired, or in noisy environments—then moment selection bias arises. In such cases, the assumption of random sampling across moments is violated. Some evidence supports this concern, as studies have found that attention-demanding activities (e.g., engaging in physical activity or being outside the home) are associated with increased likelihood of missing EMA responses (McLean, Nakamura, and Csikszentmihalyi 2017; Rintala et al. 2020).

RATIONALE FOR COMPARISON WITH TIME DIARY DATA

Time diaries are often considered the gold standard for measuring time use in large-scale surveys because of their short recall period, typically within 24 hours (Laurenceau and Bolger 2005). In a typical time diary, respondents provide a chronological account of activities from the previous day, including what they were doing, who was present, and where the activities took place. This method has helped researchers better understand the socially structured nature of daily life (Cornwell et al. 2019). The popularity of this method has led to many validation studies that examine the accuracy of time diaries. By comparing diaries with direct observation and wearable cameras and accelerometers, studies found that time diary data produce similar means for many activities (Bulungu et al. 2022; Gershuny et al. 2020; Harms

Table 1. Major Sources of Bias for Ecological Momentary Assessment and Time Diary Methods.

	Ecological Momentary Assessment	Time Diary
Data collection	High frequency (real-time or near real-time prompts several times a day)	Retrospective, typically once per day covering a 24-hour period
Sources of bias	Moment selection bias when respondents miss their responses because they are busy, tired, or in noisy environments	Recall bias, especially for brief, concurrent, or nonsalient activities

et al. 2019; Keadle et al. 2023). As such, time diaries serve as a useful benchmark for evaluating the validity of EMA data.

Recording events within 24 hours helps minimize recall bias, but time diaries are not entirely immune to this bias. For example, studies have shown that the time diary method is less accurate for concurrent or nonsalient activities (Bulungu et al. 2022; Keadle et al. 2023). These activities have lower accuracy because people are often multitasking while eating (e.g., watching TV, socializing) or caring for others (e.g., doing household chores while watching a child), which can complicate both respondent recall and how observers record the primary activity. Therefore, comparing time diary data with EMA data across different domains can offer valuable insights into how methodological choices shape reported patterns of daily activities and where discrepancies are most likely to emerge.

THE PRESENT STUDY

Numerous validation studies have examined EMA and time diary methods using objective measures such as wearable cameras and direct observation (Harms et al. 2019; Keadle et al. 2023; Stinson et al. 2022), but most have relied on small convenience samples, limiting the generalizability of their findings to broader populations. In EMA-specific studies, small sample sizes may also contribute to the wide variability in reported agreement rates. These limitations highlight the need for comparisons of EMA and time diary methods using population-based samples.

Both time diaries and EMA have emerged as strategies to reduce the recall bias commonly associated with stylized survey questions (Juster et al. 2003), although each approach has distinct strengths and limitations (see Table 1). Time diaries, although helpful in minimizing long-term recall error, remain vulnerable to bias, particularly for brief, concurrent, or nonsalient activities (Bulungu et al. 2022; Keadle et al. 2023). EMA, in contrast, significantly reduces recall error by capturing behavior in real time, but it is more susceptible to moment selection bias and issues with missing data (Stone, Schneider, and Smyth 2023). Despite these differences, few empirical studies have directly compared the estimates produced by these two methods (Klumb and

Baltes 1999). Instead, methodological preferences among social scientists often follow disciplinary norms rather than comparative evidence. This study addresses that gap by comparing time diary and EMA data drawn from two population-based samples to assess the validity, bias, and interpretive implications of each approach.

METHODS

We leverage data from two separate sources to address our research question. Both data sources provide large samples on older adults that were collected via probability sampling from the general population. In what follows, we outline the steps we took to harmonize the data, allowing us to make meaningful comparisons across the two data sources.

Social Environment and Cognitive Health in Urban and Rural Areas

EMA data come from the Social Environment and Cognitive Health in Urban and Rural Areas (SECHURA) study, a state-representative sample of Indiana residents 55 years and older. During November 2023 and March 2024, the SECHURA study conducted computer-assisted personal interview surveys with 509 respondents (a 61 percent response rate). Immediately after completing the survey, respondents were invited to participate in a follow-up EMA module that required them to use a smartphone app to record their responses. Of the 509 respondents, 272 participated in this module (53.4 percent). For the following seven days after the computer-assisted personal interview survey, EMAs were administered four random times per day, between 8 and 10 AM and between 12 and 8 PM, by "pinging" respondents through smartphone notifications. Each notification was spaced at least two hours apart. These momentary assessments aimed to capture a representative sample of the private and public spheres that respondents visited during their daily routines. Each EMA session included 8 to 14 items asking respondents to provide real-time reports about where they were (setting), what they were doing (activity), and whom they were with (social tie) and to assess their cognitive and emotional well-being. Respondents had 20 minutes to respond to each notification, with three additional reminder notifications being sent in 5-minute intervals. A vast majority of respondents (84 percent) completed their EMAs in less than 2 minutes, with an average time of 1 minute 18 seconds. The 272 respondents who participated in the EMA module produced 1,174 person-days (consisting of 5,766 EMAs).[1] We use person-days as our main unit of analysis. Sampling weights were provided by SECHURA to address sampling and response rates. Respondents who did not own smartphones ($n = 9$) were offered loaner phones to reduce selection bias. Internet access was required only during the installation phase of the LifeData app. Further details on the SECHURA research design can be found elsewhere (see Roth et al. 2024).

American Time Use Survey

The time diary data are from the American Time Use Survey (ATUS), which is conducted by the U.S. Census Bureau with funding from the Bureau of Labor Statistics to collect information on how Americans spend their time (Bureau of Labor Statistics 2025). ATUS is a nationally representative sample that draws on a subset of households from the Current Population Survey each month. An eligible person from the household (i.e., at least 15 years old) is randomly selected to be interviewed. Interviews are conducted by telephone using a computer-assisted telephone instrument. The time diary provides 24-hour recall diaries from each respondent, starting at 4 AM the previous day and ending at 4 AM on the interview day. ATUS collects information about what the respondent was doing, how long each activity lasted, where each activity occurred, and whom the respondent was with. Rather than using scripted questions, interviewers engage in conversation as an interviewing technique to obtain precise, accurate duration of activity measures. This flexible interviewing style allows interviewers to probe in a nonleading way, to guide respondents through memory lapses, and it allows respondents to describe their activities with thoroughness. The shortest unit of time reported for a given activity is five minutes, which allows for up to 288 activities on a given day, thus providing a fine-grained accounting of everyday social dynamics. Sampling weights were provided by ATUS to address sampling and response bias (Abraham, Maitland, and Bianchi 2006).

To ensure a meaningful comparison between the ATUS and SECHURA data, we applied several inclusion criteria. First, because SECHURA was conducted between 2023 and 2024, we used the 2023 ATUS data, which included 8,548 respondents (a 36.9 percent response rate). The 2024 ATUS data were not included, as they had not yet been released. Second, we restricted the ATUS sample to 2,052 respondents from midwestern states to align with the geographic focus on Indiana in SECHURA. Third, we limited the sample to respondents 55 years and older to match the SECHURA age range. These inclusion criteria resulted in an analytic sample of 1,113 ATUS respondents. Finally, we included only time diary entries from 8 to 10 AM and from 12 to 8 PM to correspond with the time frames used in SECHURA.

Measures

Social Accompaniment. Social accompaniment, which captures whether respondents are in the presence of other people or alone, is measured in similar fashion across the two samples. The ATUS measures social accompaniment using the following question: "Who was in the room with you/Who accompanied you?" Respondents were coded as alone if no one else was in the room or location where the activity took place. Emailing or phoning a friend would therefore be coded as "alone" if the friend was not physically present.

SECHURA respondents were asked, "When you heard the notification, who were you with?" Respondents were coded as alone if they selected nobody. To ensure consistency with ATUS, which counts only in-person contact as not being alone, we applied the same criterion in SECHURA. A follow-up question asked SECHURA respondents, "How were you interacting with this person(s)?" with response options of "face-to-face," "telecommunication," and "both." Respondents were recoded as being alone if they indicated their interaction was through telecommunication only.

Location. We used three broad classifications to assess the locations respondents visited during their study period: home, workplace, and the third place. In the ATUS time diary, these places were assessed using the following question: “Where were you during the activity?” An activity was coded as occurring at home if it took place in the “respondent’s home or yard” and as occurring at the workplace if it took place in the “respondent’s workplace.” All other locations were categorized as third places—spaces outside the home (the “first place”) and the workplace (the “second place”)—following Oldenburg and Brissett’s (1982) concept of public settings where people gather, socialize, and foster community.

SECHURA respondents were asked a similarly worded question: “Where were you when you heard the notification?” The response categories in SECHURA were adapted from top choices in ATUS to include “my home,” “someone else’s home,” “workplace,” “restaurant/bar/café,” “store,” “place of worship,” “library,” “park,” “somewhere else outdoors,” “healthcare facility,” and “other.” We again coded “my home” as home, “workplace” as workplace, and all other locations were coded as the third place.

Activity. Finally, we assessed the activities respondents reported engaging in throughout the study period. ATUS respondents were asked the following question for each moment of their diary day: “What were you doing?” The ATUS uses a six-digit system to code all activities. There were 17 broad categories: “personal care,” “household activities,” “caring for & helping household members,” “caring for & helping non-household members,” “work & work-related activities,” “education,” “consumer purchases,” “professional & personal care services,” “household services,” “government services & civic obligations,” “eating and drinking,” “socializing, relaxing, and leisure,” “sports, exercise, & recreation,” “religious and spiritual activities,” “volunteer activities,” “telephone calls,” and “traveling.”

SECHURA respondents were asked a similar question: “What were you doing when you heard the notification?” Their response categories were “eating/drinking,” “socializing,” “relaxing,” “working,” “shopping,” “household chores,” “volunteering,” “transporting [e.g., car, bus, bike],” “medical care,” and “other.”

This study focuses on two activity types: “household activities” and “eating and drinking.” All other activities were grouped as “other.” We selected these two categories for two key reasons: (1) they were among the most frequently reported by older participants, and (2) they represent conceptually aligned categories across ATUS and SECHURA.

Covariates. We include the following covariates in our final analysis: age (years), gender (0 = men, 1 = women), race (0 = non-White, 1 = White, because the majority of respondents in Indiana are White), education (1 = less than high school, 2 = high school or GED, 3 = some college or technical school, and 4 = college and above), partnership status (0 = not partnered, 1 = married or cohabiting), employment status (0 = not working, 1 = currently working), weekday (0 = Saturday or Sunday, 1 = Monday to Friday), and holiday (0 = no, 1 = yes).

Analytic Strategy

Each SECHURA respondent received four randomly timed EMAs per day over a 10-hour period, for seven consecutive days. This structure allows us to estimate the proportion of EMAs spent engaged in a specific activity by calculating the frequency of that activity relative to the total number of valid (nonmissing) EMAs per day. For example, if a respondent reported being at home during two of four valid EMA notifications in a day, their estimated proportion of EMAs at home would be 0.50. Assuming the prompts were randomly distributed throughout the 10-hour period, this proportion provides an approximation of the average time the individual spent at home during that window: proportion of time spent in an activity (EMA) = (frequency of this activity)/(total number of nonmissing prompts).

In contrast, the ATUS diary data provide exact durations for each activity over a 24-hour period. We restricted our analysis to the same 10-hour window used in EMA to allow an appropriate comparison. We next calculated the total time spent engaged in each activity during that window and converted this to a proportion by dividing it by the total nonmissing time recorded during the 10-hour period. For example, if a participant spent 4 hours at home during the 10-hour window, their proportion of time at home would be 0.40: proportion of time spent in an activity (time diary) = (time spent in this activity)/(total time of nonmissing activities during 10-hour window).

To assess differences between the two methods, we first estimated the proportion of time spent in each activity per day in EMA and time diary data separately. For the EMA data, we fit multilevel linear regression models to account for the nested structure of the data, as each SECHURA respondent could contribute up to seven person-day observations (see Table S1 in the online supplement). For the time diary data, we used linear regression models to predict the proportion of time spent engaged in each activity per day (see Table S2 in the online supplement). All models applied survey weights to account for sampling design and differential response rates (Abraham et al. 2006). To adjust for demographic and contextual differences between the EMA and time diary samples, we included covariates for gender, age, race, education, employment status, partnership status, whether the day was a holiday, and whether it was a weekday.

After estimating the models separately for EMA and time diary data, we stored the predicted values representing the proportion of time spent in each activity per day. We then used an indicator variable approach (Wooldridge 2016) to compare differences across data sources (see Table S3 in the online supplement). Specifically, we regressed the stored estimates on an indicator variable for data source (0 = time diary, 1 = EMA). To account for clustering in the EMA data, we used robust standard errors clustered at the respondent level. Finally, we calculated Cohen's d to evaluate the magnitude of the differences (Cohen 1988). All analyses were conducted in R and Stata. Replication code is available online at https://github.com/PengSiyun/EMA-and-Time-diary-comparison.

Table 2. Descriptive Statistics (n = 2,887).

Variable	EMA (1,774 Person-Day Observations Nested within 272 Respondents)			Time Diary (1,113 Person-Day Observations)			
	Mean	S.D.	Range	Mean	S.D.	Range	p-Value
Women	.51			.55			.459
Age	66.10	7.70	55–88	67.56	8.18	55–85	.104
White	.87			.90			.346
Education							.044
Less than HS	.05			.06			
HS or GED	.27			.39			
Some college	.24			.24			
College	.43			.31			
Employed	.40			.43			.574
Partnered	.64			.61			.587
Alone	.64	.21	0–1	.55	.36	0–1	<.001
Location							
Home	.73	.21	.06–1	.67	.31	0–1	.010
Workplace	.10	.16	0–.71	.11	.20	0–.98	.708
Third place	.17	.17	0–.89	.22	.27	0–1	<.001
Activities							
Eating/drinking	.14	.10	0–.75	.08	.07	0–.44	<.001
Household chores	.10	.10	0–.50	.18	.19	0–.95	<.001
Other activity	.76	.14	.25–1	.74	.21	0–1	.130

Note: EMA = ecological momentary assessment; HS = high school.

RESULTS

As shown in Table 2, the two samples were demographically similar in terms of gender, age, race, employment status, and partnership status. The only notable difference was in education ($p < .05$), with the EMA sample having a higher proportion of college-educated individuals. These results suggest that no major demographic differences were likely to bias time use estimates in each of the two samples.

Respondents in the EMA sample were more likely to be alone (0.64 vs. 0.55, $p < .001$), at home (0.73 vs. 0.67, $p < .01$), and eating or drinking (0.14 vs. 0.08, $p < .001$) compared with those in the time diary sample. EMA respondents were also less likely to be in the third place (0.17 vs. 0.22, $p < .001$) and performing household chores (0.10 vs. 0.18, $p < .001$). There are no significant differences across the samples in time spent at the workplace (0.10 vs. 0.11, $p = .708$) or engaging in other activities (0.76 vs. 0.74, $p = .130$). These results provide preliminary evidence of discrepancies between the two methods across various activity domains.

EMA versus Time Diary

After adjusting for demographic variables (gender, age, race, education, employment status, and partnership status) and contextual factors (whether the day was a holiday and whether it was a weekday), Figure 1 shows that respondents in the EMA sample

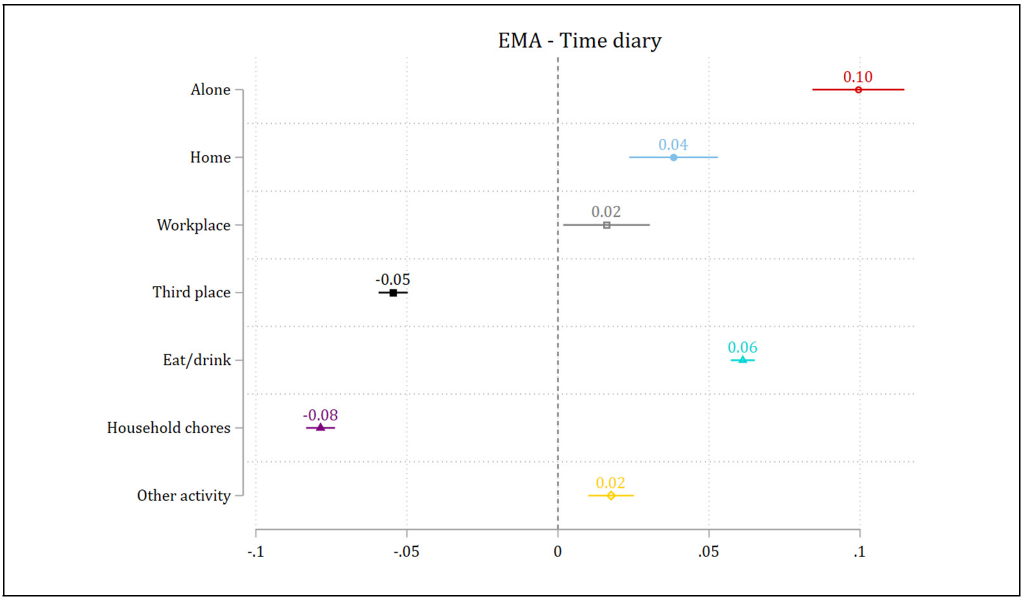

Figure 1. Differences in activity proportions between EMA and time diary (n = 2,887).
Note: For details, see Table S3 in the online supplement. EMA = ecological momentary assessment.

were more likely to be alone (b = 0.10), at home (b = 0.04), at the workplace (b = 0.02), eating or drinking (b = 0.06), and engaged in other activities (b = 0.02) compared with those in the time diary sample. EMA respondents were less likely to be in other locations (b = −0.05) and performing household chores (b = −0.08) relative to those in the time diary sample.

These patterns closely align with the unadjusted results in Table 2, suggesting that demographic and contextual differences between the two samples account for little of the discrepancy in estimates across the EMA and time diary methods. The only exceptions are the categories "workplace" and "other activity," which shift from nonsignificant to significant after adjustment. Notably, the workplace estimate changes from −0.01 (p = .708) to 0.02 ($p < .05$), indicating that after adjusting for employment status and other covariates, a statistically significant difference emerges between the two methods for time spent at work.

One challenge in validation studies is determining what constitutes a meaningful difference (Stinson et al. 2022). Because of the large sample size in this study, even small differences (e.g., 0.02) were statistically significant at the $p < .05$ level. Therefore, it is important to interpret differences using effect size. As shown in Figure 2, we converted the differences from Figure 1 into Cohen's d to assess the magnitude of the effects. According to Cohen (1988), cutoff values for small, moderate, and large effects are 0.2, 0.5, and 0.8, respectively. According to these criteria, we observed a large effect size for the following variables: being in third place (d = −1.43), eating or drinking (d = 2.11), and performing household chores (d = −1.97); a moderate effect for being alone (d = 0.77); small effects for being at

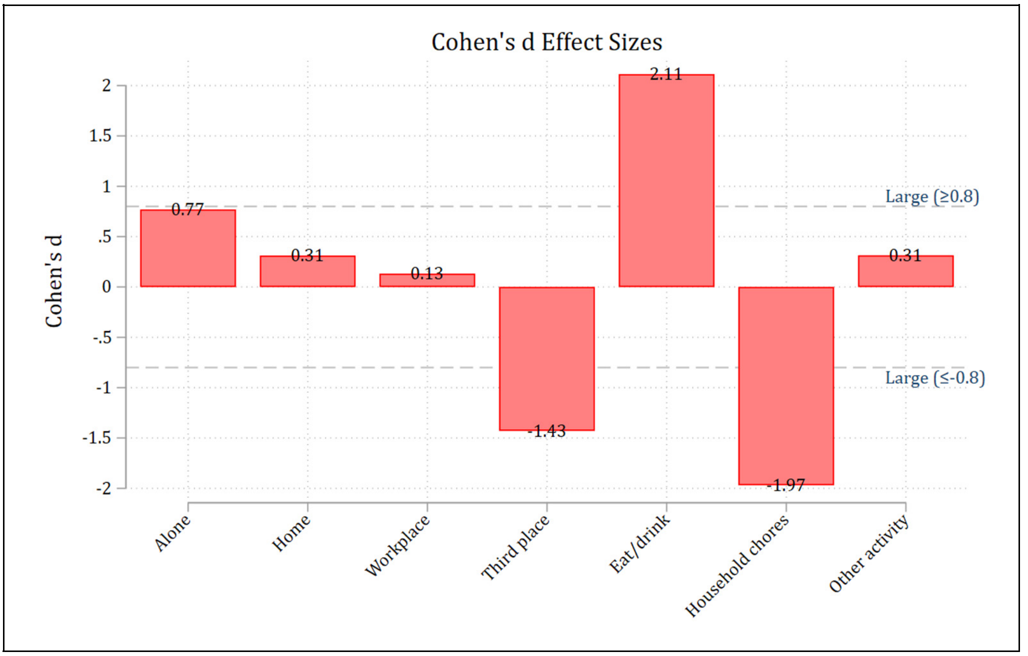

Figure 2. Cohen's *d* effect sizes for differences in activity proportions between ecological momentary assessment and time diary (n = 2,887).

Note: For Cohen's *d*, values of 0 to 0.19 indicate a negligible effect, 0.20 to 0.49 a small effect, 0.50 to 0.79 a moderate effect, and 0.80 or higher a large effect (Cohen 1988).

home (d = 0.31) and engaged in other activities (d = 0.31); and a negligible effect for time spent in the workplace (d = 0.13).

To contextualize the differences in estimates shown in Figures 1 and 2, we present the proportions of activities reported in EMA and the time diary in Figure 3. For example, being located at home and a third place show similar absolute differences of 0.04 and 0.05, respectively, in Figure 1. However, these differences translate to a small effect size for home (d = 0.31) but a large effect size for third place ($d = -1.43$). This discrepancy is explained by the baseline proportions: the proportion of activities at home was 0.73 in EMA versus 0.70 in the time diary; for third place it was 0.17 in the EMA versus 0.23 in the time diary. As this example illustrates, a similar absolute difference can yield a much larger effect size when it occurs within a smaller baseline proportion, as seen with third place.

Sensitivity Analyses

Rather than coding the proportion of activities at the person-day level in the EMA data, it is also reasonable to aggregate this measure at the person-week level. For example, if a participant reported being at home in 10 of 28 valid prompts over a week, their estimated proportion of activities at home would be 0.36. Aggregating the outcome at the respondent level simplifies the data structure by removing the nesting

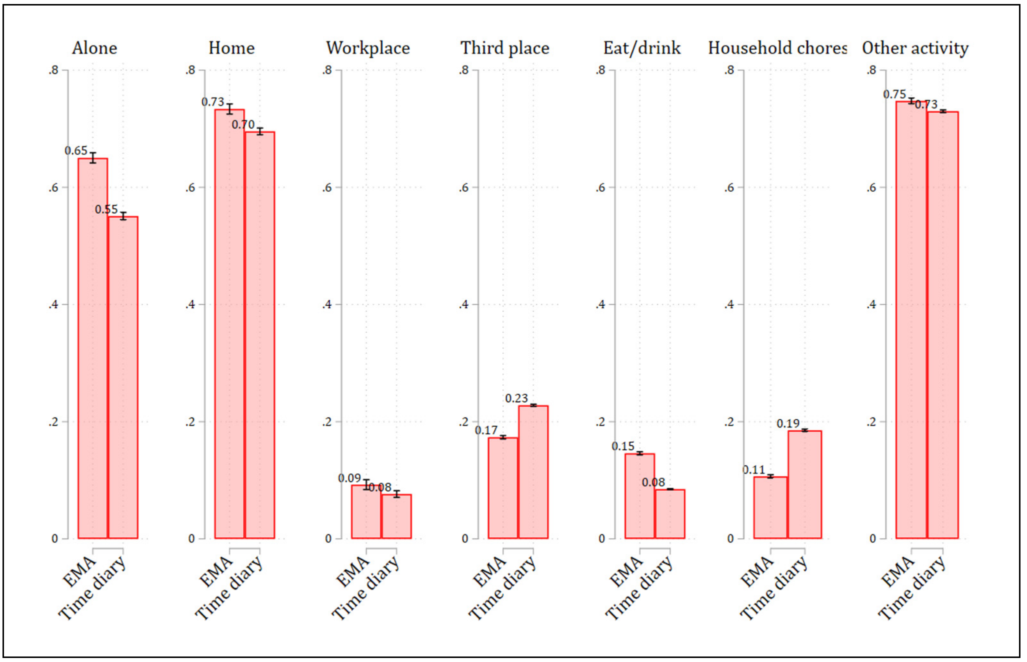

Figure 3. Proportions of activities estimated in EMA and time diary (n = 2,887).
Note: The figure includes 83.4 percent confidence intervals to facilitate visual comparison of means. Nonoverlapping intervals suggest a statistically significant difference at approximately the $p < .05$ level (Payton, Greenstone, and Schenker 2003). EMA = ecological momentary assessment.

of days within individuals, allowing a more straightforward regression approach. To compare the EMA and time diary data, we estimated linear regression models using an indicator variable for data source (0 = time diary, 1 = EMA) as the key predictor, adjusting for survey weights and demographic covariates. The results were not substantially different except for workplace and other activity (see Figure S1 in the online supplement), indicating the robustness of our findings regarding modeling strategies. Aggregating EMA outcomes to the person-week level prevents the inclusion of contextual covariates such as whether the day was a holiday or a weekday. The inability to account for these day-level factors may explain some discrepancies in estimates, particularly regarding being in the workplace. Moreover, it is more intuitive to compare EMA and time diary data at the person-day level, given the time diary data cannot be aggregated at the person-week level. For these reasons, we present the person-day level analysis as our primary approach.

As a sensitivity analysis, we restricted the time diary data to respondents from Indiana only, allowing a more direct comparison with the Indiana-based EMA data. As shown in Figure S2 in the online supplement, the results remain largely unchanged, supporting the robustness of our findings with respect to state of residence. The only exception is at the workplace, for which the difference changes from 0.02 ($p < .05$) to 0.01 ($p < .05$), becoming statistically nonsignificant. This change is not unexpected given that the effect size for this difference is negligible in the main analysis.

Given that the EMA data were collected between November 2023 and March 2024, we conducted a sensitivity analysis using time diary data from the same months to assess potential seasonal effects. As shown in Figure S3 in the online supplement, the results remain consistent, supporting the robustness of our findings with respect to seasonal variation.

Some may question whether the frequency-based measure of proportion of time in EMA is comparable with the time-based measure in time diaries. To address this concern, we constructed a frequency-based measure within the time diary by randomly drawing four time points from a 10-hour window (one between 8 and 10 AM and three between 12 and 8 PM, matching the EMA design). We then recorded the activities occurring at those time points. This randomization simulates the EMA frequency-based approach within the time diary. To minimize the influence of outliers from any single draw, we repeated the random selection 100 times and averaged the results. Finally, we compared the frequency-based and time-based measures. Figure S4 in the online supplement shows no meaningful differences, suggesting the time-based measure is comparable to the frequency-based measure.

Although our models adjust for employment status, one could argue that restricting the sample to retired or unemployed participants might further reduce nonresponse or delayed response because of workplace constraints. To address this, we conducted a sensitivity analysis limited to retired and unemployed participants. As shown in Figure S5 in the online supplement, the results remain largely unchanged, supporting the robustness of our findings to employment status.

DISCUSSION

This study leveraged two population-based samples to compare EMA and time diary data in their measurement of daily activities, examining what individuals are doing, where they are, and with whom they are interacting. The large sample size enables us to assess the magnitude of differences in estimates between the two methods. This assessment has been difficult in prior research that often relies on small and convenience samples, leading to effect size estimates with wide confidence intervals (Harms et al. 2019; Klumb and Baltes 1999; Stinson et al. 2022). As a result, many validity studies have relied on statistical significance as the primary criterion for evaluating bias, contributing to inconsistency across studies in defining and evaluating validity. For instance, studies with similarly sized differences between EMA and objective measures have drawn conflicting conclusions depending on whether the *p*-values reached statistical significance (Stinson et al. 2022). To address this limitation, we adopted effect size metrics (i.e., Cohen's *d*), which are less influenced by sample size than significance tests, to more accurately evaluate the magnitude and substantive significance of the observed differences.

Social Accompaniment

We found a moderate difference between EMA and time diary estimates for time spent alone. Specifically, the EMA method tended to overestimate moments alone compared with the time diary method. This pattern aligns with prior findings, such as Klumb and Baltes (1999), who reported a higher proportion of time alone in the EMA relative to time diaries in a sample of 83 respondents. One likely explanation for this discrepancy is moment selection bias in EMA. Indeed, respondents may be more likely to miss prompts when they are engaged in social interactions or otherwise distracted by others (Stone et al. 2023). Because social encounters are typically salient and memorable, they are less prone to recall bias in time diary methods. These findings suggest EMA may systematically underreport social interactions and overreport moments of solitude. In other words, time diaries offer more reliable estimates for aggregate measures of social interaction.

Location

We observed a small difference between EMA and time diary estimates for moments spent at home, a negligible difference at the workplace, and a large difference for moments spent in third places. The small and negligible differences observed for home and workplace settings are consistent with prior EMA validation studies, which have shown that respondents are less likely to miss EMA prompts in these locations compared with others (McLean et al. 2017; Reiter and Schoedel 2024; Rintala et al. 2020). This alignment between EMA and time diary estimates suggests that both methods yield relatively representative data for moments spent at home and in the workplace. In contrast, EMA substantially underreported exposure to third places—such as restaurants, cafés, stores, and outdoor settings—compared with time diaries. This discrepancy is consistent with prior evidence that EMA compliance decreases outside of the home and workplace (McLean et al. 2017; Reiter and Schoedel 2024; Rintala et al. 2020). We also conducted supplemental analysis and found that respondents with higher completion rates were less likely to report EMA while in third places, suggesting respondents are likely to miss EMA when in such locations (see Figure S6 in the online supplement).

Activities

We found large differences in moments spent eating or drinking and doing household chores and a small difference in other activities. EMA respondents substantially overreported moments spent eating or drinking and underreported moments spent doing household chores. This discrepancy may be due to recall bias related to concurrent activities in time diary data. Previous studies show low agreement between time diaries and objective measures for multitasking scenarios, such as eating while watching TV, drinking while socializing, or doing chores while supervising children (Bulungu et al. 2022; Keadle et al. 2023). These overlapping activities complicate respondents' recall and researchers' classification of the primary activity. Although EMA is less susceptible to recall bias for concurrent activities because of its real-time nature, it can still introduce bias when only the primary activity is recorded, as is the case in our EMA and ATUS samples. A better approach for capturing multitasking is to allow respondents to report multiple activities or explicitly prompt for secondary activities (Phipps and Vernon 2009; Rinderknecht, Doan, and Sayer 2023). The large discrepancies in eating and drinking and in household chores suggest that relying solely on primary activity reports may introduce substantial bias for activities prone to multitasking. In contrast, the small difference observed for other activities, typically less likely to occur concurrently, suggests relatively higher accuracy in those domains.

Implications

Systematic biases in EMA estimates of daily life moments have critical consequences for both statistical inference and theory development in studies that rely on these data. Because EMA missingness is not random but varies by what respondents are doing, where they are, and with whom they are interacting, systematic measurement error may be introduced that poses challenges to this method. Any research question that conditions on demographic or contextual moderators (e.g., examines whether factors like age, gender, socioeconomic status, or neighborhood disadvantage influence the association between two other variables) is especially vulnerable to error. Specifically, if those moderators are also correlated with patterns of EMA missingness (Markowski et al. 2021), researchers may unknowingly confound true behavioral or psychological differences with methodological artifacts, producing spurious associations or masking real ones.

Consider the hypothetical case of an association between solitude and alcohol use. Our results show that EMA protocols overcapture solitary moments and undercapture time spent in third places (e.g., restaurants, bars, cafés). As a result, individuals who frequently consume alcohol in social settings may be less likely to respond to prompts during those activities, whereas solitary drinking at home is more likely to be recorded. This could lead to a biased conclusion that solitude is associated with heavier alcohol use, even when individuals' overall drinking patterns remain consistent across social contexts. Similar biases could arise in studies of physical activity if active moments occur outdoors where prompts are missed, in studies of stress response if emotionally charged social interactions divert attention from the EMA prompt, or in research on the gendered division of labor if multitasked household chores are underreported.

Given these challenges, researchers should explicitly consider and attempt to mitigate the potential impact of biases in EMA estimates of daily moments on their associations of interest. For example, EMA studies could build in very brief follow-ups when EMAs are missed, asking only about important time use factors (e.g., what a person is doing, where, and with whom) to contextualize when and why EMA nonresponse occurs. Additionally, EMA missingness can be modeled as a selection bias using inverse-probability weighting, or researchers could conduct sensitivity simulation models that adjust for different missingness assumptions. Finally, EMAs could be paired with brief end-of-day surveys (when compliance with EMA prompts is high) to mimic time diary methods and report retrospectively on activities, locations, and social accompaniment that occurred during missed prompts. Without careful attention to how EMA nonresponse bias shapes data, these studies risk potentially drawing false conclusions about human behavior, cognition, and psychology. Thoughtful study design and rigorous sensitivity analyses and reporting of missingness are essential to maintain the inferential value of this highly valuable real-time method.

Limitations

This study relied on two probability samples to evaluate the difference between EMA and time diary data. Although sample differences may contribute to the observed differences, we demonstrated that samples were demographically similar, and our models adjusted for demographic and contextual factors. Additionally, we conducted sensitivity analyses to assess the robustness of our findings in light of potential sample differences.

We did not have an objective measure of daily life moments to serve as a benchmark, so we cannot definitively determine whether time diaries or EMA more accurately reflect reality. However, obtaining objective time use data, such as through direct observation or wearable cameras, is nearly impossible in a population-based sample. Despite this limitation, our study contributes to the literature by assessing the validity and potential bias of EMA and time diary data within a population-based context. The large sample size also provides greater statistical power, allowing us to examine effect sizes rather than relying solely on p-values to assess validity.

Finally, we classified social accompaniment to include only face-to-face interactions, an approach in line with the design of the ATUS. As a result, this classification may overestimate instances in which respondents were alone, particularly for people who regularly engaged in digital communication. Given that digital contact plays an important role in maintaining social connections (Peng and Roth 2022; Peng et al. 2018), future research should examine differences between EMA and time diary methods in capturing digital forms of social interaction.

CONCLUSION

By comparing data from EMA and time diaries, we evaluate the validity, potential biases, and interpretive implications of each method across various daily activities. For instance, both methods produce similar estimates for moments spent at home and in the workplace, supporting their validity in capturing activities in these settings. However, EMA tends to underestimate aggregate accounts of social accompaniment compared with time diaries, likely because of moment selection bias. Substantial discrepancies in estimates for eating/drinking and household chores indicate that relying solely on primary activity reports may lead to significant bias, particularly for activities prone to multitasking.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the National Institute on Aging (grants R01AG057739, R01AG070931, R01AG078247, and R01AG076032). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ORCID iDs

Siyun Peng https://orcid.org/0000-0002-4049-4485
Brea L. Perry https://orcid.org/0000-0001-7869-7168
Adam R. Roth https://orcid.org/0000-0002-9294-5444

Data Availability

SECHURA replication data can be requested from the corresponding author; ATUS data are publicly accessible.

Supplemental Material

Supplemental material for this article is available online.

Note

1. This yields a compliance rate of 76 percent.

References

Abraham, Katharine G., Aaron Maitland, and Suzanne M. Bianchi. 2006. "Nonresponse in the American Time Use Survey: Who Is Missing from the Data and How Much Does It Matter?"*Public Opinion Quarterly* 70(5):676–703.

Andorka, Rudolf. 1987. "Time Budgets and Their Uses." *Annual Review of Sociology* 13:149–64.

Browning, Christopher R., Catherine A. Calder, Bethany Boettner, Jake Tarrence, Kori Khan, Brian Soller, and Jodi L. Ford. 2021. "Neighborhoods, Activity Spaces, and the Span of Adolescent Exposures." *American Sociological Review* 86(2):201–33.

Browning, Christopher R., Nicolo P. Pinchak, Catherine A. Calder, and Bethany Boettner. 2024. "Leveraging Experience Sampling/Ecological Momentary Assessment for Sociological Investigations of Everyday Life." *Annual Review of Sociology* 50:41–59.

Bulungu, Andrea L. S., Luigi Palla, Jan Priebe, Lora Forsythe, Pamela Katic, Gwen Varley, Bernice D. Galinda, et al. 2022. "Validation of an Automated Wearable Camera-Based Image-Assisted Recall Method and the 24-h Recall Method for Assessing Women's Time Allocation in a Nutritionally Vulnerable Population: The Case of Rural Uganda." *Nutrients* 14(9):1833.

Bureau of Labor Statistics. 2025. "American Time Use Survey: U.S. Bureau of Labor Statistics." Retrieved January 26, 2026. https://www.bls.gov/tus/database.htm.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge.

Compernolle, Ellen L., Laura E. Finch, Louise C. Hawkley, and Kate A. Cagney. 2022. "Home Alone Together: Differential Links between Momentary Contexts and Real-Time Loneliness among Older Adults from Chicago during versus before the COVID-19 Pandemic." *Social Science & Medicine* 299: 114881.

Cornwell, Benjamin, Jonathan Gershuny, and Oriel Sullivan. 2019. "The Social Structure of Time: Emerging Trends and New Directions." *Annual Review of Sociology* 45:301–20.

Gershuny, Jonathan, Teresa Harms, Aiden Doherty, Emma Thomas, Karen Milton, Paul Kelly, and Charlie Foster. 2020. "Testing Self-Report Time-Use Diaries against Objective Instruments in Real Time." *Sociological Methodology* 50(1):318–49.

Gerstel, Naomi, and Dan Clawson. 2018. "Control over Time: Employers, Workers, and Families Shaping Work Schedules." *Annual Review of Sociology* 44:77–97.

Harms, Teresa, Jonathan Gershuny, Aiden Doherty, Emma Thomas, Karen Milton, and Charlie Foster. 2019. "A Validation Study of the Eurostat Harmonised European Time Use Study (HETUS) Diary Using Wearable Technology." *BMC Public Health* 19(2):455.

Juster, F. Thomas, Hiromi Ono, and Frank P. Stafford. 2003. "An Assessment of Alternative Measures of Time Use." *Sociological Methodology* 33(1):19–54.

Keadle, Sarah Kozey, Shreya Patel, David Berrigan, Cami N. Christopher, Jeffery Huang, Pedro F. Saint-Maurice, Erikka Loftfield, and Charles E. Matthews. 2023. "Validation of ACT24 Version 2.0 for Estimating Behavioral Domains, Active and Sedentary Time." *Medicine & Science in Sports & Exercise* 55(6):1054.

Klumb, Petra L., and Margret M. Baltes. 1999. "Validity of Retrospective Time-Use Reports in Old Age." *Applied Cognitive Psychology* 13(6):527–39.

Larson, Reed, and Maryse H. Richards. 1994. *Divergent Realities: The Emotional Lives of Mothers, Fathers, and Adolescents*. New York: Basic Books.

Laurenceau, Jean-Philippe, and Niall Bolger. 2005. "Using Diary Methods to Study Marital and Family Processes." *Journal of Family Psychology* 19(1):86–97.

Lundberg, George A., Mirra Komarovsky, and Mary Alice McInerny. 1934. "The Arts and Leisure." Pp. 253–306 in *Leisure: A Suburban Study*. New York: Columbia University Press.

Markowski, Kelly L., Jeffrey A. Smith, G. Robin Gauthier, and Sela R. Harcey. 2021. "Patterns of Missing Data with Ecological Momentary Assessment among People Who Use Drugs: Feasibility Study Using Pilot Study Data." *JMIR Formative Research* 5(9):e31421.

McLean, Derrick C., Jeanne Nakamura, and Mihaly Csikszentmihalyi. 2017. "Explaining System Missing: Missing Data and Experience Sampling Method." *Social Psychological and Personality Science* 8(4):434–41.

Monnaatsie, Malebogo, Gregore I. Mielke, Stuart J. H. Biddle, and Tracy L. Kolbe-Alexander. 2024. "Ecological Momentary Assessment of Physical Activity and Sedentary Behaviour in Shift Workers and Non-shift Workers: Validation Study." *Journal of Sports Sciences* 42(10):874–83.

Noh, Jung Min, SongHyun Im, JooYong Park, Jae Myung Kim, Miyoung Lee, and Ji-Yeob Choi. 2025. "Validation of Ecological Momentary Assessment with Reference to Accelerometer Data: Repeated-Measures Panel Study with Multilevel Modeling." *Journal of Medical Internet Research* 27(1):e59878.

Oldenburg, Ramon, and Dennis Brissett. 1982. "The Third Place." *Qualitative Sociology* 5(4):265–84.

Overton, Mark, Sarah Ward, Nicola Swain, Carrie Falling, David Gwynne-Jones, Roger Fillingim, and Ramakrishnan Mani. 2023. "Are Ecological Momentary Assessments of Pain Valid and Reliable? A Systematic Review and Meta-Analysis." *Clinical Journal of Pain* 39(1):29.

Payton, Mark E., Matthew H. Greenstone, and Nathaniel Schenker. 2003. "Overlapping Confidence Intervals or Standard Error Intervals: What Do They Mean in Terms of Statistical Significance?" *Journal of Insect Science* 3(1):34.

Peng, Siyun, and Adam R. Roth. 2022. "Social Isolation and Loneliness Before and During the COVID-19 Pandemic: A Longitudinal Study of U.S. Adults Older Than 50." *Journals of Gerontology: Series B* 77(7):e185–90.

Peng, Siyun, Merril Silverstein, J. Jill Suitor, Megan Gilligan, Woosang Hwang, Sangbo Nam, and Brianna Routh. 2018. "Use of Communication Technology to Maintain Intergenerational Contact: Toward an Understanding of 'Digital Solidarity.'" Pp. 159–80 in *Connecting Families? Information & Communication Technologies in a Life Course Perspective*, edited by B. B. Neves and C. Casimiro. Bristol, UK: Policy Press.

Phipps, Polly A., and Margaret K. Vernon. 2009. "Twenty-Four Hours: An Overview of the Recall Diary Method and Data Quality in the American Time Use Survey." In *Calendar and Time Diary Methods in Life Course Research*, edited by R. F. Belli, F. P. Stafford, and D. F. Alwin. Thousand Oaks, CA: Sage.

Reiter, Thomas, and Ramona Schoedel. 2024. "Never Miss a Beep: Using Mobile Sensing to Investigate (Non-)Compliance in Experience Sampling Studies." *Behavior Research Methods* 56(4):4038–60.

Rinderknecht, Robert G., Long Doan, and Liana C. Sayer. 2023. "Secondary Activities: Their Proximity to Primary Activities and Their Importance for Understanding Reports of Preparing and Consuming Meals." *Survey Methods: Insights from the Field*. Retrieved January 26, 2026. https://surveyinsights.org/?p=17465.

Rintala, Aki, Martien Wampers, Inez Myin-Germeys, and Wolfgang Viechtbauer. 2020. "Momentary Predictors of Compliance in Studies Using the Experience Sampling Method." *Psychiatry Research* 286:112896.

Roth, Adam R. 2024. "Ecological Momentary Assessments in Sociology." *Social Currents* 11(2):103–11.

Roth, Adam R., Siyun Peng, Jafnun Nudrat Jarin, Tianyao Qu, Maleah Fekete, and Brea L. Perry. 2024. "Social Environment and Cognitive Health in Urban and Rural Areas (SECHURA)." Retrieved January 26, 2026. https://sciety.org/articles/activity/10.31235/osf.io/9n5pz?utm_source=sciety_labs_article_page.

Solhan, Marika B., Timothy J. Trull, Seungmin Jahng, and Phillip K. Wood. 2009. "Clinical Assessment of Affective Instability: Comparing EMA Indices, Questionnaire Reports, and Retrospective Recall." *Psychological Assessment* 21(3):425–36.

Stinson, Lesleigh, Yunchao Liu, and Jesse Dallery. 2022. "Ecological Momentary Assessment: A Systematic Review of Validity Research." *Perspectives on Behavior Science* 45(2):469–93.

Stone, Arthur A., Stefan Schneider, and Joshua M. Smyth. 2023. "Evaluation of Pressing Issues in Ecological Momentary Assessment." *Annual Review of Clinical Psychology* 19:107–31.

Sullivan, Gail M., and Richard Feinn. 2012. "Using Effect Size—or Why the *P* Value Is Not Enough." *Journal of Graduate Medical Education* 4(3):279–82.

Wooldridge, Jeffrey M. 2016. *Introductory Econometrics: A Modern Approach*. 6th ed. Boston: Cengage Learning.

Wrzus, Cornelia, and Andreas B. Neubauer. 2023. "Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance across Research Fields." *Assessment* 30(3):825–46.

Author Biographies

Siyun Peng is an assistant professor in the School of Aging Studies at the University of South Florida. His research focuses on aging, health, social networks, and quantitative methods. He brings a broad interdisciplinary background in sociology, gerontology, and statistics, with expertise in applying a life-course perspective to investigate the social determinants of health.

Brea L. Perry is the Allen D. and Polly S. Grimshaw Professor in the Department of Sociology at Indiana University. Her areas of research include social networks, biosociology, social inequalities, aging, medical sociology, and mental health. She was a National Academy of Medicine Emerging Leaders in Health and Medicine Scholar (2019–2022) and received the 2025 Leo G. Reeder Award for career contributions to medical sociology from the American Sociological Association. Her work has been funded by the National Institutes of Health, the National Science Foundation, and charitable foundations.

Adam R. Roth is an assistant professor at Oklahoma State University. His work appears in outlets such as *Social Networks, Social Science & Medicine, Sociology of Health & Illness, and Journals of Gerontology: Social Sciences*. He is the principal investigator of the SECHURA study.