# Integrating Sensor and Survey Data: Representativeness in Event-based Sampling

**Mareike Wieland[1], Charlotte de Alwis[1], Lukas Otto[1], Julian Kohne[1]**
**[1] GESIS – Leibniz Institute for the Social Sciences, Cologne**

Mobile devices enable the passive collection of behavioral data, capturing both digital traces (e.g., app usage logs) and contextual information (e.g., geo-location). In the social sciences, such data are often complemented by surveys to explain behavior through individual traits and situational factors. Traditionally, this involves cross-sectional and retrospective self-reports linked to behavioral data. However, smartphones also allow for innovative, time-sensitive survey designs. Intensive longitudinal methods (such as experience sampling or momentary assessments) already common in psychology are increasingly used in social science research to collect data multiple times per day. In so called event-based designs, passive data trigger in-situ self-reports, enabling close synchronization of behavior and subjective experience while capturing intra-individual variation (Barret & Barret, 2001; Schnauber-Stockmann & Karnowski, 2020).

## Event-based Designs: Potentials and Challenges

There are several approaches to implement event-based designs. Using manual logging, study participants recognize a relevant event themselves and submit an event-related self-report (Otto et al., 2021). To reduce biases due to self-selection, randomized prompts can actively remind participants to log an event if it occurred in the previous time interval. The most advanced approach uses automated detection of events within behavioural data as real-time triggers for immediate self-reports (e.g., Exler et al., 2018), allowing researchers to synchronize experience and behaviour with minimal delay (Masur, 2019). Such automated event-based approaches to integrating behavioral and survey data in situ thus offer significant, yet largely untapped, potential for the social sciences: (1) They reduce recall bias and memory-related errors by prompting participants immediately after relevant events. (2) They enable the situational assessment of both the subjective meaning of events and context-specific individual characteristics, allowing temporal variation to be modeled as a source of insight rather than error. (3) They align the temporal resolution of both data types by linking a large amount of real-time self-reports to continuous behavioral data.

Despite these advantages, automated event-based designs are rarely implemented. Their limited adoption can be attributed to their complexity in all steps along the research process, from conceptualization and design to technical implementation and data analysis. One of the key challenges central to this presentation is selecting the right sampling strategy – on both the participant and the event level. On the participant level, the willingness to participate in passive data collection studies is already widely discussed, with privacy concerns and technical barriers or lack of digital literacy identified as major obstacles (e.g., Keusch et al., 2019), making recruiting a sufficiently large and suitable sample challenging. On the level of events, special attention must be paid to the timing of survey prompts. Since these prompts are distributed several times a day, they can easily be perceived as overly burdensome and time-consuming, leading to survey fatigue, reduced compliance, or dropout (Stone et al, 2021).

This is especially true for automated event-based prompting, where a third and largely underexplored challenge arises. Researchers typically lack prior knowledge of the distribution of relevant events, making it impossible to draw a random sample of event occurrences. Attempting to capture every event through self-report prompts is equally problematic, as the number of prompts might be overwhelming. To address this challenge, researchers must rely on criteria-based, multi-level sampling strategies that aim to approximate representativeness while balancing the variability of target behaviours and the cognitive and temporal demands placed on participants. Yet, such sampling protocols have rarely been subjected to systematic empirical evaluation. To address this gap, we use a case study and examine:

*RQ: How well does a criteria-based multi-level sampling protocol for self-report prompts following automatic event detection capture the full range of relevant behavioural events?*

**Case Study**

Data were collected as part of a larger study on social media use and political information among German social media app users aged 18 to 54 (Wieland, 2024). *App usage* was passively recorded over one week using the commercial tool murmuras. To enable intensive longitudinal and, specifically, *event-based automatic self-report prompting*, the Android-only tool was customized in accordance with the event sampling protocol outlined in the subsequent paragraph. Participants were recruited in March 2021 in cooperation with the panel provider Norstat, using soft quotas for education, age, and gender. The analysis includes only those participants for whom both app logging and in situ self-report data are available. The final dataset comprises 71,788 app usage events and 12,498 completed self-report prompts from 376 participants.

***Event sampling protocol***

The case study examined how users perceive news encounters during social media use, with a focus on how these perceptions relate to characteristics of the usage episode. Based on the general and information-related relevance of specific platforms, usage episodes involving Facebook, Instagram, X (Twitter), and YouTube are defined as the population of relevant events. Prior usage studies suggest that such episodes occur frequently throughout the day (Deng et al., 2019), with short "checking" sessions more common in the morning and longer, more focused use often taking place in the early evening. Additionally, the sequential use of multiple platforms is expected to emerge as a habit over time (Bayer & LaRose, 2018). These dynamics argue against simple prompting protocols, such as setting a fixed maximum number of prompts starting early in the day. Such approaches would likely fail to capture the temporal and behavioral variance of actual usage, thereby compromising the ecological validity of the findings. Instead, a multi-level, situational sampling protocol was defined and implemented through triggering rules in the data collection app. Table 1 summarizes the criteria and underlying rationale guiding the automatic event sampling.

*Table 1 Event-based sampling protocol*

| Rule | Criteria | Rationale |
|---|---|---|
| Post-Episode Prompting | Surveys are triggered after app usage (when the app moves to the background). | Prevents interruptions during use and reduces the likelihood of behavioural reactivity to the survey content. |
| Minimum Interaction Threshold | App usage must exceed 15 seconds. | Ensures that only episodes involving meaningful engagement are sampled (vs. accidental or extremely brief app openings; see also Lukoff et al., 2018) |
| Time-Based Distribution Control | A maximum of 4 surveys per 3-hour window, starting from the first usage event of the day. | Prevents clustering of data in specific time slots (e.g., morning only), promotes coverage of diverse daily contexts, and increases the chance of capturing variation in social media usage behaviour. (see also Stone et al., 2021) |
| Minimum Time Between Surveys | A 30-minute minimum interval between completed surveys. | Avoids excessive interruptions, participant fatigue, and potential reactivity. |

### *Analysis Strategy*

The quality of social science data analyses largely depends on the quality of the underlying sample. Setting aside the participant level, the following analysis investigates how well the criteria-based sampling protocol outlined in Table 1 captures the actual population of study-relevant events. To do so, we compare the distribution of app usage episodes across the four selected social media platforms with the distribution of submitted responses to event-based prompts for self-reports.

At the participant level, we examine whether the sample (Q1) captured all social media apps used by each participant and (Q2) how accurate it reflects the relative frequency with which different platforms were used. At the aggregate level, we investigate whether the sampling protocol adequately captured social media usage events (Q3) across different times of day, (Q4) on different weekdays, and (Q5) across the study period. Finally, we assess (Q6) whether the distribution of captured events reflects the actual variation in event duration.

**Results**

To assess whether all social media apps used by a participant were covered by prompts, we first examined which and how many of the four selected social media platforms each participant used during the period of the study. App logging data show that approximately 20% of participants used only one of the four social media platforms during the study week. Around 39% used two, 34% used three, and only 8% used all four apps. Accordingly, the prompting protocol should ensure that all apps used by a participant are also represented in their self-reports. This objective was largely met: for 82% of participants, the submitted ESM surveys covered all social media apps used, indicating that the sampling protocol provided comprehensive coverage (Q1).

We further examined how accurately the sampling protocol reflects platform-specific usage patterns (Q2). Figure 1 shows that, overall, 17 percent of all logged social media app usage episodes triggered a successful prompt for in situ self-reporting. While usage episodes on Facebook and Instagram are covered close to the average, YouTube episodes are captured at a higher rate, and Twitter at a lower rate.

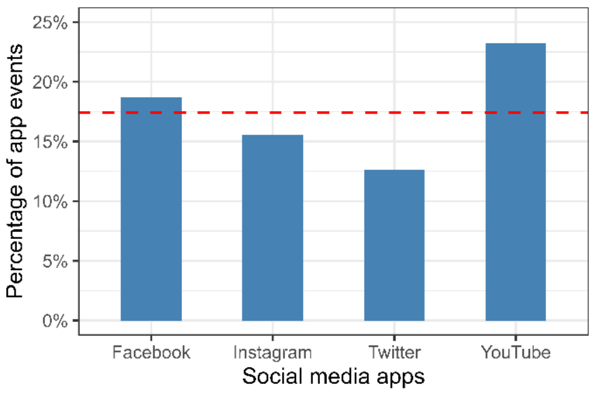


*Figure 1 – Share of app events captured through self-report prompts by social media app.*

This discrepancy may partly be explained by differences in baseline event frequency: YouTube, Facebook, and Instagram were each used by roughly two-thirds of participants (Fig. 2), whereas only 19 percent used Twitter, resulting in fewer opportunities to trigger prompts. Additionally, differences in average session length—YouTube having the longest—may contribute, as participants might be more inclined to respond to prompts during longer usage episodes (Fig. 3): Since our analysis only includes successful prompts (i.e., submitted self-reports), higher compliance to submit a self-report on usage experiences during longer sessions may skew the coverage in favor of platforms like YouTube. This underlines the complexity of designing a criteria-based protocol for sampling events in advance. Without knowing the actual parameters, such as the usage frequency and duration of different social media apps, the usage of different platforms is only partly matched by the sampling protocol (Q2).

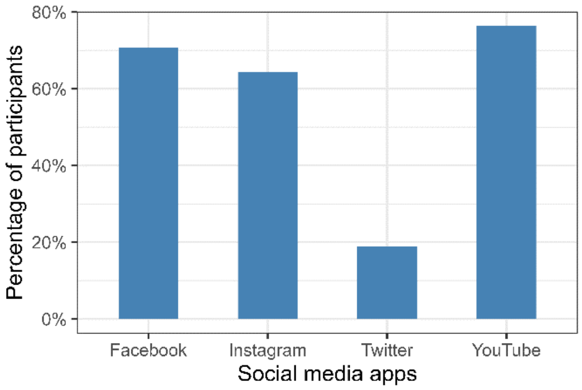


*Figure 2 – Share of participants using each social media app.*

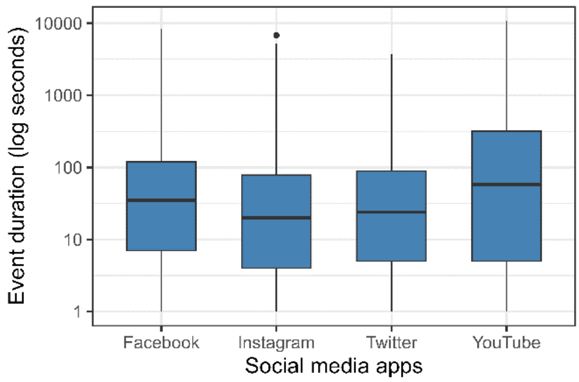


*Figure 3 – Duration of app usage events by social media app (log-transformed seconds[1]).*

Furthermore, we examined whether the sampling protocol resulted in a balanced rate of prompts throughout the day. Figure 4 shows that app usage and corresponding prompts are distributed almost identically over time. Figure 5 confirms this with a nearly constant proportion of events captured by prompts. Thus, the sampling protocol provided good coverage of app usage through prompts (Q3).

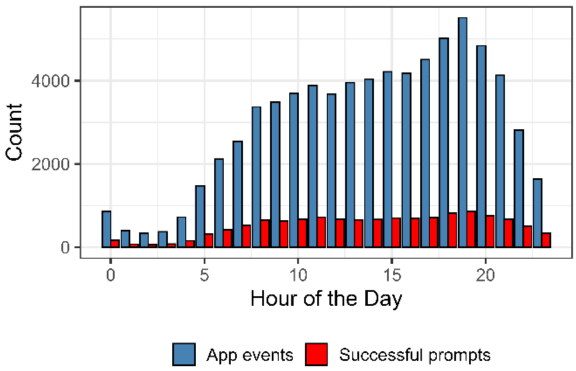


*Figure 4– Number of app events and successful self-report prompts by hour of the day.*

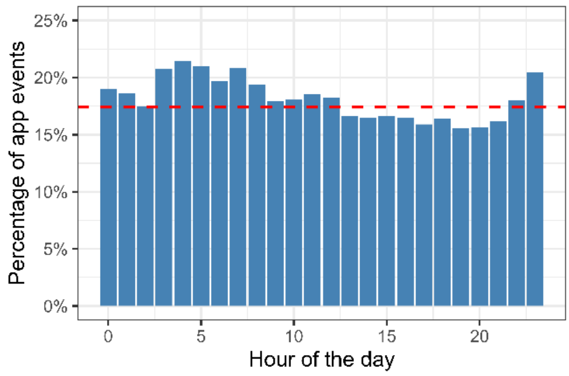


*Figure 5 – Share of app events captured through self-report prompts by hour of the day.*

Regarding coverage across different weekdays, no remarkable deviations were observed. The study day, which differs from the weekday due to participants' staggered start dates, shows slight

[1] Log scale (base 10) applied to y-axis for readability.

variations on the share of app events captured by self-report prompts on days two and three. Nonetheless, for both Q4 and Q5, the protocol maintained a consistent coverage of events through prompts.

Finally, we examined prompt coverage based on the session length of app usage, ranging from 1 second up to almost three hours. Figure 6 shows that longer usage episodes triggered survey prompts more often (almost 25%) than the overall average (17%). In contrast, shorter sessions are clearly underrepresented, with less than 10 percent resulting in prompts. The sampling protocol is thus not optimal when it comes to an accurate representation of app usage session with different length in self-report prompts especially in edge cases (Q6).

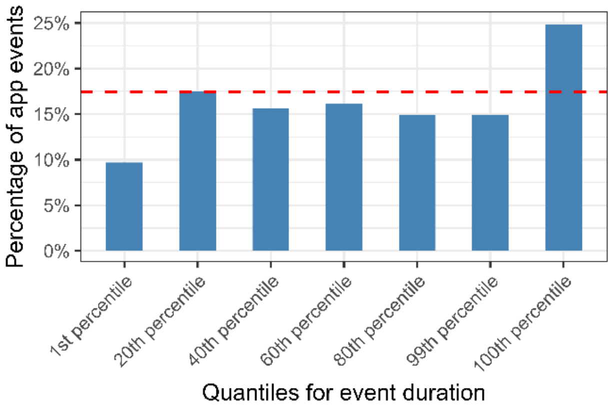


*Figure 6 - Share of app events captured by ESMs by event duration*

## Discussion

In summary, the sampling protocol resulted in a good coverage of social media app usage across most evaluated dimensions through event-based self-reports. Dividing the day into three-hour intervals proved particularly effective, resulting in an almost identical distribution of app usage and triggered prompts throughout the day. The sampling protocol performs somewhat less well when multiple parameters interact in ways that are difficult to anticipate in advance. Our findings indicate that, in particular, the duration of app usage systematically interacts with the protocol, such that longer usage episodes as well as apps that are on average used for longer periods are more likely to trigger self-report prompts than shorter ones and are consequently overrepresented in the sample. However, these findings must be interpreted with caution, as the current analysis is limited to successful prompts only. As a result, the effects of the sampling protocol might be confounded with issues of participant compliance. Although the overall average response rate (submitted self-reports/all prompted self-reports) was 64% across all study days, indicating acceptable compliance, it remains unclear whether shorter episodes were systematically excluded due to the protocol itself or because participants were less inclined to provide self-reports for brief usage episodes. Since this is an artifact of the data provision by the tool provider, this limitation underscores the clear need for carefully designed technical solutions that prioritize data quality and comprehensive metadata.

To address these and further issues, we plan to extend the analysis to the situational level to understand better which types of events are over- or underrepresented and whether the sampling protocol equally captures app usage patterns of different participant groups.

## References


Barrett, L. F., & Barrett, D. J. (2001). An Introduction to Computerized Experience Sampling in Psychology. *Social Science Computer Review*, *19*(2), 175-185. https://doi.org/10.1177/089443930101900204

Bayer, J. B. & LaRose, R. (2018). Technology Habits: Progress, Problems, and Prospects. In B. Verplanken (Hrsg.), The Psychology of Habit. Theory, Mechanisms, Change, and Contexts (S. 111–130). Cham: Springer International Publishing.

Deng, T., Kanthawala, S., Meng, J., Peng, W., Kononova, A., Hao, Q., Zhang, Q. & David, P. (2019). Measuring smartphone usage and task switching with log tracking and self-reports. Mobile Media & Communication, 7(1), 3–23. https://doi.org/10.1177/2050157918761491

Exler, A., Kramer, S., Meza Martínez, M.A., Navolskyi, C., Vogt, M., Beigl, M. (2018). Suitability of Event-Based Prompts in Experience Sampling Studies Focusing on Location Changes. In: Perego, P., Rahmani, A., TaheriNejad, N. (eds) Wireless Mobile Communication and Healthcare. MobiHealth 2017. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 247. Springer, Cham. https://doi.org/10.1007/978-3-319-98551-0_19

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly, 83*(S1), 210–235. https://doi.org/10.1093/poq/nfz007

Lukoff, K., Yu, C., Kientz, J. & Hiniker, A. (2018). What Makes Smartphone Use Meaningful or Meaningless? Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(1), 1–26. https://doi.org/10.1145/3191754

Masur, P. K. (2019). Capturing situational dynamics: Strength and pitfalls of the experience sampling method. In P. Müller, S. Geiß, C. Schemer, T. Naab, & C. Peter (Eds.), *Dynamische Prozesse der öffentlichen Kommunikation – Methodische Herausforderungen.* Herbert von Halem Verlag.

Otto, L. P., Thomas, F., Glogger, I., & De Vreese, C. H. (2021). Linking Media Content and Survey Data in a Dynamic and Digital Media Environment – Mobile Longitudinal Linkage Analysis. Digital Journalism, 10(1), 200–215. https://doi.org/10.1080/21670811.2021.1890169

Schnauber-Stockmann, A., & Karnowski, V. (2020). Mobile Devices as Tools for Media and Communication Research: A Scoping Review on Collecting Self-report Data in Repeated Measurement Designs. *Communication Methods and Measures*, *14*(3), 145–164. https://doi.org/10.1080/19312458.2020.1784402

Stone, A. A., Obbarius, A., Junghaenel, D. U., Wen, C. K. F. & Schneider, S. (2021). High-resolution, field approaches for assessing pain: Ecological Momentary Assessment. Pain, 162(1), 4–9. https://doi.org/10.1097/j.pain.0000000000002049

Wieland, M. (2024). *Informiert oder (doch nur) abgelenkt?.* Herbert von Halem Verlag. https://doi.org/10.1453/2023_9783869623863